

RELATÓRIOS COPPEAD

429

Setembro 2016

MODELAGEM
DE RISCO
FINANCEIRO

Beatriz Vaz de Melo
Mendes

Relatórios COPPEAD é uma publicação do Instituto COPPEAD de Administração da Universidade Federal do Rio de Janeiro (UFRJ)

Editora

Leticia Casotti

Editoração

Lucilia Silva

Ficha Catalográfica

Cláudia de Gois dos Santos

Vaz de Melo Mendes, Beatriz.

Modelagem do risco financeiro. / Beatriz Vaz de Melo Mendes. – Rio de Janeiro: UFRJ/COPPEAD, 2016.

234 p.; 27 cm. – (Relatórios COPPEAD; 429)

ISSN 1518-3335

ISBN 978-85-7508-116-7

Modelagem do Risco Financeiro

Beatriz Vaz de Melo Mendes

Instituto de Matemática/COPPEAD - UFRJ

Rio de Janeiro, 11 de maio de 2015

Prefácio

Este livro é uma consequência natural das notas de aula utilizadas em diversos cursos de graduação e mestrado do IM/UFRJ, COPPEAD/UFRJ, e do IMPA. Os alunos demandavam uma versão mais organizada e atualizada dessas notas já que o material coberto somente poderia ser encontrado em algum conjunto de títulos estrangeiros e nacionais. Contudo, nenhum deles isoladamente continha todos os tópicos abordados e/ou não possuía nosso enfoque prático. Víamos a necessidade de várias modificações com o objetivo de atualizar e adaptar o conteúdo ao interesse dos alunos. Este texto é resultado dessas modificações.

Assim, este texto aborda, de uma maneira abrangente e atual, o problema da análise de dados, em particular da análise de séries temporais e da mensuração do risco financeiro. A exposição inclui desde as técnicas baseadas em normalidade ou simplesmente não paramétricas, até as modelagens dinâmicas e sofisticadas conhecidas hoje, como por exemplo aquelas baseadas em estimação robusta e na abordagem via pair-copulas condicionais. Alguns tópicos mais importantes dos nossos textos "Introdução a Análise de Eventos Extremos" de 2004, e "Riscos Financeiros" de 2007, foram revistos e atualizados aqui. Acrescentamos alguns conceitos estatísticos necessários para o bom entendimento das análises feitas, e organizamos as versões atualizadas dos programas utilizados nas análises em SPlus (e em R). Em resumo, nesta nova versão das notas de aula apresentamos o estado da arte para a modelagem de ativos financeiros e para o cálculo do risco, considerando também os aspectos dinâmicos das modelagens e o tratamento de eventos extremos.

Seu nível é introdutório e o enfoque é aplicado. Ênfase é dada na interpretação dos conceitos mostrando como as metodologias funcionam na prática. Existem excelentes livros com maior rigor matemático, os quais são citados ao longo do texto, mas nos parece que tanto o estudante, seja de graduação ou pós-graduação, quanto o profissional irá se beneficiar dessa leitura bastante concisa que pode ter sua força justamente na sua utilidade prática. O nível é de graduação, e o leitor precisa ter boa noção de conceitos de probabilidade e inferência neste nível.

Em termos de estruturação, este livro está dividido em sete capítulos. O Capítulo 1 mostra como fazer uma análise exploratória básica e rotineira de dados, fornecendo as ferramentas necessárias para a primeira etapa de sua análise estatística. O Capítulo 2 trata da modelagem univariada não condicional de retornos financeiros, revendo as distribuições de probabilidade mais utilizadas e/ou adequadas para a modelagem da distribuição subjacente das séries. O Capítulo 3 fornece a versão multivariada do Capítulo 2. Nele modelamos a distribuição conjunta de d variáveis através de distribuições elípticas multivariadas, cópulas e pair-copulas. O Capítulo 4 trata da modelagem da média condicional de séries

temporais, e no Capítulo 5 estudamos modelos para a volatilidade condicional. No Capítulo 6 vemos a versão multivariada do Capítulo 4, inclusive Cointegração. No Capítulo 7, focamos em medidas de risco baseadas nos modelos vistos ao longo do texto. Pela sua generalidade e flexibilidade, essas técnicas podem ser aplicadas a dados oriundos de outras áreas, já que risco é um conceito abrangente que também aparece em Epidemiologia, Meio Ambiente, Atuária, etc. Em todos os capítulos utilizamos dados reais para ilustrar as modelagens introduzidas e sugerimos alguns exercícios. Finalmente, o Apêndice fornece as referências bibliográficas.

Agradecemos a todos, ex-alunos, professores, familiares, que nos apoiaram e contribuíram na elaboração dessas notas. Vários alunos da graduação e pós-graduação do Instituto de Matemática e do Instituto COPPEAD da UFRJ, e do IMPA, colaboraram de diferentes formas com a elaboração deste texto ao longo dos últimos 6 anos. A eles meu agradecimento e meu carinho.

Finalmente, a autora agradece àqueles que a incentivaram a escrever essas notas, seus filhos queridos, José Hugo, Paula e Júlia.

Beatriz Vaz de Melo Mendes

Rio de Janeiro, 11 de maio de 2015

Sumário

1	Propriedades Empíricas de Retornos Financeiros	11
1.1	Retornos Financeiros	12
1.2	Análises Estatísticas Básicas de Retornos Financeiros	16
1.3	Fatos Estilizados de Séries de Retornos	21
1.4	Correlações e Dependências entre Extremos	23
1.5	Apêndice do Capítulo 1	25
1.6	Exercícios	26
2	Modelagem Não Condicional Univariada	29
2.1	Distribuição Normal	31
2.2	Distribuição Log-Normal	33
2.3	Distribuições Estáveis	33
2.4	Mistura de Normais	34
2.5	Distribuição Normal Assimétrica	35
2.6	Distribuições Elípticas	36
2.7	Distribuição t -Student	41
2.8	Distribuições Elípticas Assimétricas	42
2.9	Distribuição t Assimétrica	43
2.10	Valores Extremos	45
2.11	Apêndice do Capítulo 2	69
2.12	Exercícios	72
3	Modelagem Não Condicional Multivariada	75
3.1	Cópulas	76
3.2	Pair-copulas	82
4	Modelagem da Média Condicional	87
4.1	Processos estocásticos, ergodicidade e estacionariedade	87
4.2	Exemplos de processos estocásticos	90
4.3	Processos lineares	93
4.4	Modelagem AR(F)IMA	95

4.4.1	Processos Autorregressivos $AR(p)$	96
4.4.2	Processos Médias Móveis $MA(q)$	106
4.4.3	Processos $ARMA(p, q)$	113
4.4.4	Processos $ARIMA(p, d, q)$	117
4.4.5	Modelos Sazonais	119
4.4.6	Processos $ARFIMA(p, d, q)$	120
4.5	Processos Não Estacionários	126
4.6	Testes de Raiz Unitária	127
4.7	Exercícios	132
5	Volatilidade	133
5.1	Fatos estilizados sobre a volatilidade condicional	134
5.2	Modelos GARCH	137
5.3	Modelos de volatilidade estocástica com memória longa	149
5.4	Ilustração	150
5.5	Volatilidade Realizada	156
5.6	Exercícios	162
6	Modelos Condicionais Multivariados	165
6.1	Modelos Vetoriais Autorregressivos ($VAR(p)$)	165
6.2	Cointegração	168
6.2.1	O modelo VECM	169
6.2.2	Testes para cointegração	172
6.2.3	Regressão espúria	175
6.3	Cointegração com limiar	175
6.4	Ilustração	180
6.5	Cópulas dinâmicas	187
6.6	Exercícios	189
6.7	Apêndice do Capítulo 6	190
7	Medidas de Risco	195
7.1	Valor em Risco	196
7.2	VaR Não Condicional	197
7.3	VaR Condicional	204
7.4	Shortfall Condicional	211
7.5	Drawdown-em-Risco e Máximo Drawdown-em-Risco	212
7.6	Qual a melhor estimativa do VaR?	214
7.7	Exercícios	215

Lista de Figuras

1.1	<i>Preços diários de fechamento dos índices Brasil, México, Indonésia, e China de 03/01/2006 a 13/10/2011.</i>	13
1.2	<i>Retornos percentuais R_t e r_t para o índice brasileiro.</i>	14
1.3	<i>A a.c.f. amostral dos log-retornos dos índices Brasil e México.</i>	20
1.4	<i>Gráfico no tempo dos log-retornos dos índices do Brasil e do México.</i>	22
1.5	<i>Na parte superior temos 30 observações da Normal bivariada (X, Y) com $\rho = 0.7$ e seus postos. Na parte inferior temos as transformações $\exp(2 * X)$ e $\exp(Y)$ e seus postos.</i>	24
2.1	<i>Histograma e densidade da Normal baseada nos estimadores de máxima verossimilhança para os índices do Brasil (esquerda) e do México (direita).</i>	32
2.2	<i>Densidade da Normal assimétrica para $\lambda = 0$ em preto, $\lambda = \pm 1$ em azul, e $\lambda = \pm 4$ em rosa.</i>	35
2.3	<i>Densidade da t assimétrica com $\mu = 0$, $\sigma = 1$, $\nu = 4$, e $\lambda = -0.6, 0.0, 0.6$, respectivamente representadas pelas linhas sólida, pontilhada, e tracejada.</i>	44
2.4	<i>Índice Brasil de 4 de janeiro de 2006 a 4 de julho de 2008.</i>	46
2.5	<i>Densidades dos três tipos de distribuições de valores extremos nas suas formas padrões. Gumbel em preto, Weibull em rosa e Fréchet em azul.</i>	48
2.6	<i>Esquerda: Densidade f da f.d.a $(1 - e^{-x})^{30}$ (linha contínua) e densidade Gumbel g (linha pontilhada) para $\mu = 3.4$ e $\sigma = 0.94$. Direita: Diferença $f - g$.</i>	49
2.7	<i>À esquerda a densidade Fréchet estimada e o histograma dos valores absolutos dos mínimos da China. À direita, o QQ-plot dos resíduos versus a distribuição de referência Exponencial(1).</i>	56
2.8	<i>Estimativas do índice extremal para várias opções de tamanho N de bloco e limiares, e para a cauda esquerda do índice do Brasil.</i>	57
2.9	<i>O tempo médio de espera versus o limiar u para os mínimos mensais do Brasil e da China.</i>	59
2.10	<i>(Esquerda) Densidades: GPD (Beta), tracejada e GEV (Weibull), contínua, ambas com parâmetro $\xi = -0, 3$. (Direita) Densidades: GPD (Pareto), tracejada com parâmetros $\xi = 1$, $\mu = -0, 25$ e $\sigma = 1/\pi$ e Cauchy padrão, contínua.</i>	62
2.11	<i>Densidades: Exponencial (pontilhada); Pareto $\xi = 0, 5$ (contínua); Beta $\xi = -0.2$ (tracejada). As três densidades possuem locação zero e escala um.</i>	63
2.12	<i>Gráfico de k versus estimativas do estimador de Hill correspondente.</i>	67

2.13	<i>Sensibilidade das estimativas por máxima verossimilhança de ξ em relação à escolha do limiar u, para a cauda esquerda do índice Brasil. No eixo horizontal o número de excessos usados no processo de estimação.</i>	68
3.1	<i>Valores simulados de uma cópula Gaussiana e uma t-student (4gl) com $\rho = 0.0$.</i>	82
3.2	<i>Representação gráfica de uma pair-copula 6-dimensional tipo D-vine.</i>	85
4.1	<i>Algumas trajetórias de um processo estocástico.</i>	88
4.2	<i>Dois processos estocásticos com variância igual a 3. À esquerda um puramente aleatório; à direita um ruído branco.</i>	91
4.3	<i>Um passeio aleatório a partir de um processo i.i.d. com distribuição t-student(3).</i>	92
4.4	<i>Trajetoárias de dois processos $AR(1)$ com $\phi_0 = 0$, $\phi_1 = 0.05$, $\phi_1 = 0.95$, a_t normal com $\sigma^2 = 0.09$ e $T=100$.</i>	98
4.5	<i>As f.a.c.'s de processos $AR(1)$ com $\phi_1 = 0.5$ do lado esquerdo e $\phi_1 = -0.5$ do lado direito.</i>	98
4.6	<i>Análise gráfica do ajuste de um $AR(1)$ aos retornos do índice brasileiro.</i>	100
4.7	<i>f.a.c.s de um $MA(1)$ com $\theta_1 = 0.25$ e de uma $MA(2)$ com $\theta_1 = 0.25$ e $\theta_2 = -0.25$.</i>	109
4.8	<i>f.a.c.'s amostrais das séries de retornos do índice da Indonésia e do México.</i>	110
4.9	<i>Série de retornos e resíduos do ajuste $MA(2)$ aos log-retornos do México.</i>	111
4.10	<i>Previsões baseadas no ajuste $MA(2)$ para os retornos do México.</i>	113
4.11	<i>Funções de autocorrelação e de autocorrelação parcial amostrais para os log-retornos diários do Brasil.</i>	118
4.12	<i>Série diferenciada da série componente sazonal de período 12.</i>	120
4.13	<i>Série diferenciada de lag 12 da série anterior.</i>	121
4.14	<i>f.a.c. de séries e $I(1)$ à esquerda e $I(0)$ ou $I(d)$, d fracionário, à direita.</i>	123
4.15	<i>Periodograma da série log-retornos da China à esquerda, e de sua diferenciação, à direita.</i>	125
5.1	<i>Preços diários de fechamento e log-retornos percentuais do índice Brasil de 03/01/2006 a 13/10/2011.</i>	150
5.2	<i>A f.a.c. amostral dos log-retornos percentuais e seus quadrados para o índice Brasil de 03/01/2006 a 13/10/2011.</i>	151
5.3	<i>As volatilidades estimadas a partir dos três modelos. A linha preta corresponde ao $GARCH(1,1)$. A rosa ao modelo com memória longa, a qual parece produzir os maiores valores para a volatilidade. Na parte superior a diária, e na parte inferior a anualizada.</i>	153
5.4	<i>As persistências a frente $\theta_{t+k t}$ e acumulada $\phi_{t+k t}$ estimadas a partir dos dois modelos sem memória longa $M1$ (esquerda) e $M2$ (direita). Em cada gráfico a linha preta corresponde a persistência para a frente, e a linha azul representa a persistência acumulada. As curvas estão multiplicadas por um fator.</i>	154
5.5	<i>Previsões da volatilidade anualizada para 200 passos a frente. Uma das datas de início das previsões corresponde a um período de baixa volatilidade (27 de janeiro de 2001) e a outra em um período de alta volatilidade (31 de outubro de 2008). À esquerda e direita os resultados para os modelos $M1$ e $M2$.</i>	155

5.6	<i>Previsões anualizadas da volatilidade 1 passo a frente e 63 dias e para os dois modelos (esquerda e direita).</i>	156
6.1	<i>Duas séries cointegradas. Séries do CDS e Volatilidade Implícita da firma AIG, no tempo e uma versus a outra.</i>	178
6.2	<i>Séries do CDS (em baixo), Volatilidade Implícita (alto) e VIX (meio).</i>	181
6.3	<i>Os erros de desequilíbrio para o modelo clássico linear.</i>	184
6.4	<i>Limiares versus SQR. Linhas verticais representam os percentuais dos resíduos ordenados definindo os possíveis limiares.</i>	186
6.5	<i>SQR ao longo do tempo. Bola azul indica data do SQR mínimo.</i>	187

Capítulo 1

Propriedades Empíricas de Retornos Financeiros

Os avanços teóricos na modelagem de dados econômico-financeiros observados nas duas últimas décadas, aliados à uma percepção mais clara de risco pelos participantes de mercados cada vez mais integrados, têm levado as instituições a aumentarem seu foco no gerenciamento de risco.

Séries financeiras possuem características especiais que tornam totalmente inadequada a aplicação de suposições simplificadoras ou matematicamente convenientes, tais como a da distribuição normal multivariada. Atenção especial é necessária com a modelagem das caudas das distribuições de dados financeiros, já que a forma dessas caudas tem relação com a ocorrência de extremos em tempos de crise. Assim, a administração de riscos financeiros deve envolver modelagens que utilizem distribuições com caudas pesadas e levem em consideração os valores extremos, mudanças de regime e interdependências complexas.

A principal preocupação dos administradores de risco é com os eventos de baixa probabilidade que possam resultar em perdas catastróficas para suas instituições e seus clientes. Por exemplo, o *crash* dos mercados financeiros em outubro de 1987 deu origem a inúmeras investigações científicas. Muitas delas concluíram que a grande perda verificada era passível de ser mensurada, através de uma probabilidade não nula, e que não teria relação com nenhuma circunstância anormal, tal como alguma derivada de fatores institucionais, mas que teria sido esta queda simplesmente uma resposta normal a um choque em um mundo de mercados agregados. Isso significa que grandes quedas em mercados financeiros não estão necessariamente associadas a notícias ou acontecimentos de grande impacto, e podem ser modeladas usando séries históricas. É interessante notar que os métodos tradicionais baseados, por exemplo, em normalidade, irão falhar quando mais se precisa deles: nos momentos de crise, quando se observam grandes movimentos de mercado.

A estimação acurada de uma medida de risco passa necessariamente por alguma suposição sobre a (ou algum ajuste de uma) distribuição dos (aos) dados. Assim, iremos nos capítulos 2 e 3 tratar

do problema da modelagem não condicional univariada e multivariada de retornos financeiros, e nos capítulos seguintes iremos considerar a dinâmica temporal dessas séries para finalmente utilizar todo este ferramental para fazer previsões e estimar medidas de risco. Ao longo deste texto iremos ilustrar os conceitos e as modelagens estatísticas utilizando um conjunto de quatro índices de mercados emergentes, Brasil, México, Indonésia e China. Essas séries foram obtidas no endereço www.msci.com e cobrem o período de 03/01/2006 a 13/10/2011. Os *scripts* utilizados nas análises são fornecidos no Capítulo 8, na linguagem “S” e na linguagem “R” no livro anexo.

1.1 Retornos Financeiros

Seja P_t o preço de fechamento de um ativo financeiro no tempo t . A escala temporal na qual os dados financeiros são coletados é muito importante pois pode enfatizar (ou o contrário) algumas características dinâmicas dos mesmos. Ela pode ser anual, semanal, diária, horária, de alta frequência, etc. Séries de preços diários em geral apresentam tendência crescente e *não* são *estacionárias* violando por isto a premissa básica para os modelos para séries temporais a serem utilizados nos capítulos seguintes.

Um processo estocástico é fracamente estacionário, ou estacionário de segunda ordem¹, se e somente se os seus primeiro e segundo momentos são invariantes no tempo. Isto é, sendo $\{X_t\}$, $t \in \mathcal{T}$ este processo, temos que $E[X_t] = \mu$, $E[X_t^2] < \infty$, constantes para todo t , e a covariância entre X_{t_1} e X_{t_2} depende apenas do *lag* $|t_1 - t_2|$, para quaisquer índices t , t_1 , e t_2 em um conjunto de índices \mathcal{T} . A Figura 1.1 ilustra este comportamento não estacionário dos preços, no caso dos índices do Brasil, México, Indonésia e China. Observamos os efeitos da crise de refinanciamento americana, crise esta que culminou com a queda de instituições financeiras importantes em setembro de 2008.

Por outro lado, retornos financeiros são em geral fracamente estacionários e por possuírem propriedades estatísticas mais interessantes, são preferidos para análise. Esta suposição pode ser verificada através de testes, por exemplo o teste KPSS de Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. e Shin, Y. (1992) dado na próxima seção, ou simplesmente dividindo a série em sub-amostras e comparando as estimativas dos dois primeiros momentos obtidas para as sub-amostras.

Diferentemente dos P_t , os retornos não dependem da unidade dos dados. Supondo que não haja dividendos pagos no período sob análise, a variação do preço P_t entre os instantes $t - 1$ e t é dada por

$$\Delta P_t = P_t - P_{t-1}.$$

O *retorno líquido simples de 1-período* deste ativo entre os instantes $t - 1$ e t é definido como

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{\Delta P_t}{P_{t-1}}.$$

Da definição acima obtemos $1 + R_t = \frac{P_t}{P_{t-1}}$, o *retorno bruto de 1-período*. É comum expressar R_t em percentagem, daí a denominação de *taxa de retorno*.

¹Um processo $\{X_t\}_{t \in \mathcal{Z}}$ é estritamente estacionário se para todo $t_1, \dots, t_n, k \in \mathcal{Z}$ e para todo $n \in \mathcal{N}$, os vetores aleatórios $(X_{t_1}, \dots, X_{t_n})$ e $(X_{t_1+k}, \dots, X_{t_n+k})$ possuem a mesma distribuição.

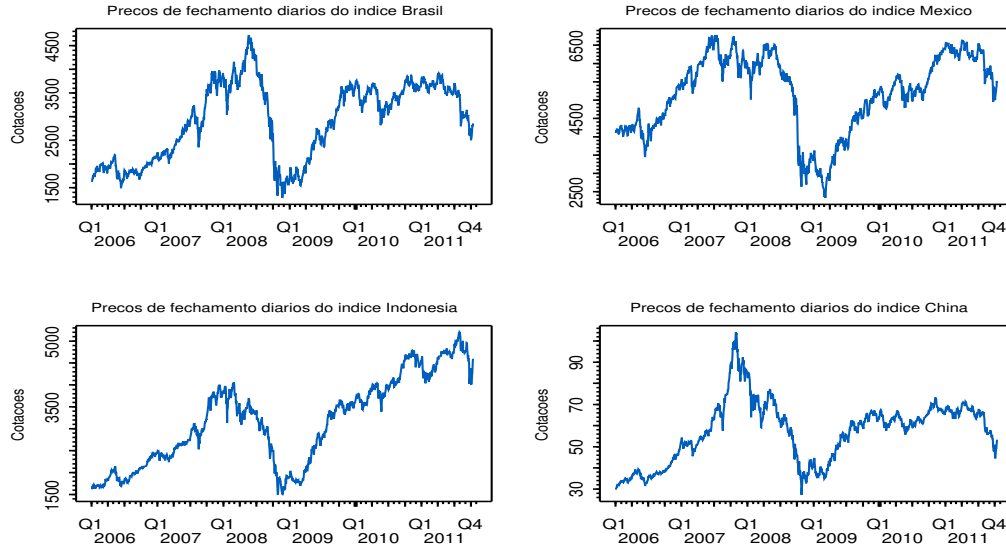


Figura 1.1: Preços diários de fechamento dos índices Brasil, México, Indonésia, e China de 03/01/2006 a 13/10/2011.

Chamamos de *retorno composto continuamente* (ver efeito da composição contínua em Tsay, R. S. (2005)) ou *log-retorno* o

$$r_t = \log \frac{P_t}{P_{t-1}} = \Delta \log P_t,$$

onde $\log(\cdot)$ representa o logaritmo neperiano, e portanto $\frac{P_t}{P_{t-1}} = e^{r_t}$.

Seja $p_t = \log P_t$. Então $r_t = p_t - p_{t-1}$. Da definição anterior temos que $r_t = \log(1 + R_t)$. Sabendo que para u pequeno, $\log(1 + u) \approx u$, temos que os valores de R_t e r_t são em geral próximos. A Figura 1.2 ilustra este fato e mostra os retornos percentuais R_t e r_t para o índice brasileiro. Daqui para frente quando nos referirmos a “retorno” estaremos nos referindo ao “ r_t ”, ou mais precisamente a $r_t = (p_t - p_{t-1}) * 100$, seu valor percentual.

Considere os tempos $t - k$ e t . O *retorno líquido simples de k -períodos* é dado por

$$R_t[k] = \frac{P_t - P_{t-k}}{P_{t-k}}.$$

Notemos que a expressão acima nos dá:

$$\begin{aligned} 1 + R_t[k] &= \frac{P_t}{P_{t-k}} = \frac{P_t}{P_{t-1}} \frac{P_{t-1}}{P_{t-2}} \frac{P_{t-2}}{P_{t-3}} \dots \frac{P_{t-k+1}}{P_{t-k}} \\ &= (1 + R_t)(1 + R_{t-1})(1 + R_{t-2}) \dots (1 + R_{t-k+1}) \end{aligned}$$

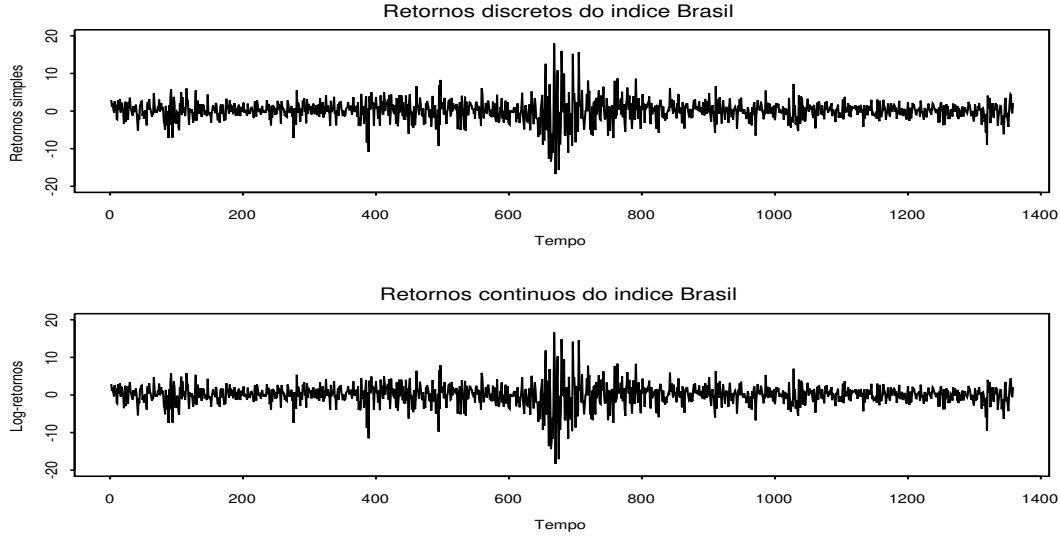


Figura 1.2: Retornos percentuais R_t e r_t para o índice brasileiro.

e vemos que o *retorno bruto de k -períodos*, $1 + R_t[k]$, é o produto dos k retornos simples de 1-período. Os retornos “anualizados” são bastante utilizados na prática, já que facilitam as comparações. São definidos como

$$R_t[k](\text{anualizado}) = [\prod_{j=0}^{k-1} (1 + R_{t-j})]^{1/k} - 1 ,$$

que é a média geométrica dos k retornos brutos de 1-período. Este valor pode ser aproximado por $(1/k) \sum_{j=0}^{k-1} R_{t-j}$, a partir da expansão de Taylor até primeira ordem.

Temos também que o *retorno composto continuamente de k -períodos*, ou o *log-retorno de k -períodos*, dado por

$$r_t[k] = \log \frac{P_t}{P_{t-k}}$$

e portanto

$$r_t[k] = \log\left(\frac{P_t}{P_{t-1}} \frac{P_{t-1}}{P_{t-2}} \cdots \frac{P_{t-k+1}}{P_{t-k}}\right) = \sum_{j=0}^{k-1} r_{t-j}$$

valendo também que

$$r_t[k] = \log((1 + R_t)(1 + R_{t-1}) \cdots (1 + R_{t-k+1})) = \sum_{j=0}^{k-1} \log(1 + R_{t-j}).$$

Sendo o retorno de k períodos soma de variáveis aleatórias, temos que a distribuição de $r_t[k]$ tenderá para a Normal pelo Teorema Central do Limite (TCL).

Podemos também definir o retorno de uma carteira. Seja C uma carteira com d componentes (retornos de ativos financeiros). Seja w_i o peso correspondente ao ativo i . O retorno simples desta carteira no tempo t é

$$R_{C,t} = \sum_{i=1}^d w_i R_{it}$$

onde $R_{it} = \frac{P_{it} - P_{i,t-1}}{P_{i,t-1}}$ é o retorno simples do ativo i . Assumimos $\sum w_i = 1$. Notemos que desta definição temos que

$$\begin{aligned} 1 + R_{C,t} &= 1 + \sum_{i=1}^d w_i R_{it} \\ &= \sum w_i + \sum_{i=1}^d w_i R_{it} = \sum w_i (1 + R_{it}). \end{aligned}$$

Logo

$$\frac{P_{C,t}}{P_{C,t-1}} = \sum_{i=1}^d w_i \frac{P_{i,t}}{P_{i,t-1}}$$

de onde obtemos

$$e^{r_{C,t}} = \sum_{i=1}^d w_i e^{r_{it}}$$

ou

$$r_{C,t} = \log\left(\sum_{i=1}^d w_i e^{r_{it}}\right).$$

Ou seja, os retornos logarítmicos de uma carteira não possuem aquela propriedade aditiva interessante que os retornos simples possuem. Contudo, quando os r_{it} são pequenos temos que $r_{C,t} \approx \sum_{i=1}^d w_i r_{it}$. Esta aproximação é frequentemente utilizada.

Podemos definir também o retorno em excesso ou *excesso de retorno*. Considere um ativo de referência, R_0 , em geral livre de risco, por exemplo, letras do Tesouro Americano de curto prazo. Então, o *retorno simples em excesso* é $Z_t = R_t - R_{0t}$. O *log-excesso de retorno* é $z_t = r_t - r_{0t}$, onde r_{0t} é o log-retorno do ativo de referência.

1.2 Análises Estatísticas Básicas de Retornos Financeiros

Nesta seção revemos alguns conceitos básicos necessários para uma análise preliminar, exploratória, de um conjunto de dados, em particular de uma série de retornos financeiros. Esses conceitos buscam caracterizar a distribuição subjacente e a relação entre as variáveis no tempo.

Antes de mais nada é preciso testar se a série é estacionária. Veja no Capítulo 4 as definições formais de estacionariedade e de ruído branco (RB). Para entender melhor as questões envolvidas, considere uma série $\{y_t\}$ cujo processo gerador é:

$$y_t = (c + \delta t) + x_t$$

$$x_t = \phi x_{t-1} + \epsilon_t, \quad \epsilon_t \sim \text{RB}(0, \sigma^2).$$

Conforme veremos também no Capítulo 4, a série x_t acima é um processo AR(1). Se $|\phi| < 1$ então y_t é estacionária em torno da tendência determinística $(c + \delta t)$. Se $\phi = 1$ então $x_t = x_{t-1} + \epsilon_t = x_0 + \sum_{j=1}^t \epsilon_j$ é uma tendência estocástica (x_t é não estacionário de raiz unitária) e portanto y_t segue uma tendência determinística linear $(c + \delta t)$ mas não necessariamente reverte para ela. Reversão para a tendência existirá no caso de y_t ser estacionária. Notemos ainda que podemos ter c ou δ ou ambos iguais ou não a zero.

Testes de raiz unitária têm hipótese nula $H_0 : \phi = 1$ (a primeira diferença seria estacionária) contra a hipótese alternativa $H_1 : \phi < 1$ (estacionária em torno de uma tendência).

Testes de estacionariedade têm hipótese nula que y_t é estacionária em torno de uma tendência. Se tomamos a primeira diferença de y_t obtemos

$$\Delta y_t = \delta + \Delta x_t$$

$$\Delta x_t = \phi \Delta x_{t-1} + \epsilon_t - \epsilon_{t-1},$$

e temos na segunda equação um modelo ARMA com raiz unitária na parte MA ($\epsilon_t - \epsilon_{t-1}$), portanto um MA não invertível. Este fenômeno é conhecido como *super diferenciação*. Assim, o teste KPSS testa a hipótese nula de que a série é estacionária, ou $I(0)$ (ver definição de $I(0)$ no Capítulo 4), testando se a série diferenciada possui uma raiz unitária na sua representação MA. É um teste unilateral para a direita. A estatística teste do teste KPSS não possui uma distribuição padrão conhecida, a qual também depende da especificação da parte determinística, e seus valores críticos são em geral obtidos por simulação.

Tendo estabelecido que a série é estacionária podemos investigar suas características não condicionais.

Seja r uma variável aleatória (v.a.) representando o log-retorno, e F a sua função de distribuição acumulada (f.d.a.), isto é, $F(x) = P(r \leq x)$, e escrevemos $r \sim F$. Supomos ser r uma v.a. contínua e seja f sua função de densidade. Denotaremos por $\mu = E[r] = \int_{-\infty}^{+\infty} r f(r) dr$ o valor esperado de r , ou *média* de r . O k -ésimo momento central de r é dado por $E[(r - \mu)^k]$. O segundo momento central é a *variância* de r , que denotaremos por σ^2 . Sua raiz quadrada, σ , é o *desvio padrão* de r . Bastam os 2 primeiros momentos para completamente definir a distribuição Normal, o que é também verdade para qualquer distribuição elíptica, conhecida sua função geradora. Outras distribuições podem necessitar do conhecimento de seus momentos de ordens maiores para sua identificação.

O terceiro momento central mede a simetria de r em relação à sua média, e o quarto momento central mede o peso da cauda de r . O *coeficiente de assimetria* de r , $A(r)$, é dado por

$$A(r) = \frac{E[(r - \mu)^3]}{\sigma^3}$$

e temos $A = 0$ para distribuições simétricas. O *coeficiente de curtose* de r , $C(r)$, é definido como

$$C(r) = \frac{E[(r - \mu)^4]}{\sigma^4}.$$

A quantidade $(C(r) - 3)$ é chamada de *excesso de curtose*, e é zero para a distribuição Normal. Uma distribuição com $C(r) - 3 > 0$ tem excesso de curtose positivo e possui caudas mais grossas que as de uma distribuição Normal.

Considere agora a sequência $\{r_t\}_{t=1}^T$, onde r_t é o log-retorno observado no instante t do tempo, em geral t sendo um dia útil de mercado. Por enquanto não faremos nenhuma suposição sobre o processo gerador desta sequência de v.a's.

Os estimadores mais utilizados para estimar μ , σ^2 , $A(\cdot)$ e $C(\cdot)$ são as suas versões amostrais, ou estimadores empíricos (veja Bickel, P.J. e Doksum, K.A. (2001)). Sejam $\bar{r} = \frac{1}{T} \sum_{t=1}^T r_t$ a *média amostral* e $S^2 = \frac{1}{T-1} \sum (r_t - \bar{r})^2$ a *variância amostral* de (r_1, r_2, \dots, r_T) . Esses são os estimadores clássicos mais utilizados para estimar a localização e a variância de F , e são também os estimadores de máxima verossimilhança² sob normalidade, isto é, assumindo-se $F = N(\mu, \sigma^2)$. Os estimadores empíricos para os coeficientes de assimetria e curtose são dados pelos coeficientes de *assimetria amostral* e de *curtose amostral*, definidos como

$$\widehat{A(r)} = \frac{1}{(T-1)S^3} \sum_{t=1}^T (r_t - \bar{r})^3,$$

onde $S = \sqrt{S^2}$ é o *desvio padrão amostral*, e

$$\widehat{C(r)} = \frac{1}{(T-1)S^4} \sum_{t=1}^T (r_t - \bar{r})^4.$$

Para várias aplicações em finanças é importante se detectar se a série de retornos possui distribuição subjacente assimétrica. Por exemplo, sendo $\{r_t\}$ estacionária, é natural que os investidores prefiram uma série positivamente assimétrica e, de fato, vários modelos em finanças têm estendido sua definição e incorporado esta característica. Como exemplos temos o *Capital Asset Pricing Model*, CAPM, e a fórmula de Black Scholes, a qual apresenta maior precisão no apreçamento de opções quando são feitas correções envolvendo a assimetria e a curtose. Vale a pena notar que testes de assimetria baseados no coeficiente de assimetria amostral devem ser avaliados com cautela devido à não normalidade dos retornos. Em Peiró, A. (1999) séries de retornos são analisadas com respeito à assimetria.

²Definição de estimadores de máxima verossimilhança: Considere uma sequência de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.), (r_1, \dots, r_T) , e seja F_θ a distribuição hipotetizada indexada por um vetor de parâmetros θ . Suponha que f represente a função de probabilidade. A função $L(\theta; r_1, \dots, r_T) = \prod_{i=1}^T f(\theta; r_i)$ é a função de verossimilhança. Os estimadores de máxima verossimilhança (EMV) de θ são os valores $\hat{\theta} \in \Theta$, Θ o espaço paramétrico, que maximizam $L(\cdot)$ ou alternativamente o $\log(L(\cdot))$.

Como parte dessas análises exploratórias básicas, podemos também efetuar alguns testes de hipóteses. O teste Jarque-Bera é bastante utilizado para testar se a distribuição subjacente dos dados seria a Normal. Se $r \sim N(\mu, \sigma^2)$, então $\widehat{A(r)}$ e $\widehat{C(r)}$ convergem em distribuição para uma Normal (Snedecor, G.W. e Cochran, W.G., 1989). Mais precisamente, quando $T \rightarrow \infty$,

$$\widehat{A(r)} \rightarrow^d N(0, \frac{6}{T}),$$

e

$$\widehat{C(r)} \rightarrow^d N(3, \frac{24}{T}),$$

onde a notação \rightarrow^d significa *converge em distribuição*.

Se de fato a distribuição dos retornos for a Normal, teremos todos os momentos centrais de ordem ímpar k , $k = 2l + 1$, $l \geq 1$, iguais a zero. Podemos testar

$$H_0 : A(r) = 0$$

versus a hipótese alternativa $H_1 : A(r) \neq 0$ usando o valor absoluto da estatística teste $\widehat{A(r)}\sqrt{\frac{T}{6}}$, a qual converge para a $N(0, 1)$.

Para testar

$$H_0 : C(r) = 3$$

em um teste bilateral usamos o valor absoluto da estatística teste $(\widehat{C(r)} - 3)\sqrt{\frac{T}{24}}$, a qual também converge para a $N(0, 1)$.

O teste de Jarque-Bera (Jarque, C.M. e Bera, A.K., 1980) combina os dois testes acima e propõe a estatística JB :

$$JB(r) = (\frac{T}{6})\widehat{A(r)}^2 + (\frac{T}{24})(\widehat{C(r)} - 3)^2,$$

a qual, sob a hipótese nula de normalidade dos retornos, tem distribuição *chi-quadrado* com 2 graus de liberdade, χ_2^2 . No caso dos log-retornos diários dos quatro índices o teste JB rejeitou a hipótese nula de normalidade com p-valor *zero* para todos os índices.

Na Tabela 1.1 vemos as estimativas de algumas estatísticas básicas no caso dos log-retornos diários dos quatro índices utilizados, a primeira etapa em qualquer análise exploratória. Extremos de maior magnitude ocorreram para o par Brasil e México, apesar de diferirem em relação à maior perda. Por outro lado, de acordo com o estimador robusto, o Brasil apresentou um ganho mediano bem maior que os demais, embora as médias amostrais sejam bastante próximas. O teste de assimetria, baseado na estatística $\widehat{A(r)}\sqrt{\frac{T}{6}}$, rejeitou fortemente (com p-valor zero) a hipótese nula de uma distribuição simétrica para o índice do Brasil, mas aceitou para os outros três índices. Além disto, o maior desvio padrão e o maior excesso de curtose também ocorreram para o índice brasileiro.

A covariância γ_k entre as v.a.'s r_{t-k} e r_t , definida como

$$\gamma_k = cov(r_{t-k}, r_t) = E[r_{t-k}r_t] - E[r_{t-k}]E[r_t]$$

para $t, k \in \mathcal{Z}$, $\mathcal{Z} = \{0, \pm 1, \pm 2, \dots\}$, é chamada de autocovariância de *lag* k de r_t . Num processo estacionário, a autocovariância depende apenas da defasagem k das variáveis no tempo e $cov(r_{t-k}, r_t) = cov(r_t, r_{t+k})$ para todo $t \in \mathcal{Z}$. Vale também que $\gamma_0 = var(r_t)$.

Tabela 1.1: Estatísticas simples das séries de log-retornos diários do Brasil, México, Indonésia e China.

Índice	Média	DP	\hat{A}	\hat{C}^*	Máximo	Mínimo	Mediana
Brasil	0.0417	2.7954	-0.4478	6.8233	16.6188	-18.3232	0.1873
México	0.0214	2.1405	0.0625	6.7024	15.1586	-10.8985	0.1243
Indonésia	0.0770	2.0517	-0.0856	6.7962	13.5995	-12.1719	0.1037
China	0.0431	2.2539	0.0069	5.2936	13.8338	-12.8377	0.0454

Convenções utilizadas na tabela: DP: desvio padrão amostral; \hat{A} : coeficiente de assimetria amostral; \hat{C}^* : excesso de curtose amostral.

O coeficiente de autocorrelação de lag k de r_t , ρ_k , é definido como $\rho_k = \frac{cov(r_t, r_{t-k})}{(var(r_t)var(r_{t-k}))^{1/2}}$. Sob a suposição de estacionariedade fraca de r_t temos que

$$\rho_k = \frac{\gamma_k}{\gamma_0},$$

para todo $k \in \mathcal{Z}$, e temos $\rho_k = \rho_{-k}$, $-1 \leq \rho_k \leq 1$, e $\rho_0 = 1$. É interessante observar os ρ_k como função do lag k , o que dá origem à *função de autocorrelação*, a qual denotaremos por f.a.c.

Uma série estacionária é *não auto-correlacionada* se e somente se $\rho_k = 0$ para todo $k > 0$. O fato de $\rho = 0$ não implica em independência. Independência entre duas variáveis r_t e r_{t+k} ocorre se a distribuição conjunta de r_t e r_{t+k} for a Normal bivariada com $\rho = 0$, ou se a cópula que liga as duas distribuições marginais é a cópula produto (neste caso as marginais podem ser qualquer F).

Dada uma série de tamanho T de retornos, a *autocorrelação amostral* de lag k de r_t , $\hat{\rho}_k$, é dada por

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}, \quad k \in \mathcal{Z},$$

onde $\hat{\gamma}_k = \frac{1}{T} \sum_{t=1}^{T-k} (r_t - \bar{r})(r_{t+k} - \bar{r})$. Se a série $\{r_t\}$ é estacionária e ergódica³ com $var(r_t) < \infty$, temos que

$$\hat{\gamma}_k \rightarrow_{q.c.} \gamma_k \quad e \quad \hat{\rho}_k \rightarrow_{q.c.} \rho_k$$

onde $\rightarrow_{q.c.}$ representa *convergência quase certa*. Se o processo $\{r_t\}$ for uma sequência *puramente aleatória*, isto é, uma sequência de v.a.'s i.i.d., e se $E[r_t^2] < \infty$, então $\hat{\rho}_1$ será assintoticamente Normal com média zero e variância $1/T$. Pode-se testar se $\rho_1 = 0$, versus $\rho_1 \neq 0$ usando a estatística $\sqrt{T}\hat{\rho}_1$ a qual assintoticamente tem distribuição $N(0, 1)$.

Em geral, para qualquer inteiro fixo k , se $\{r_t\}$ for uma sequência i.i.d. com segundo momento finito, temos que $\hat{\rho}_k$ é assintoticamente Normal com média zero e variância $1/T$. Esse resultado pode ser utilizado para se testar $\rho_k = 0$ versus $\rho_k \neq 0$, e para se construir intervalos de confiança. A estatística teste é $\sqrt{T}\hat{\rho}_1$, a qual é assintoticamente Normal. Quando $\{r_t\}$ é um processo linear com inovações i.i.d. Normais temos a fórmula de Bartlett: $\hat{\rho}_k \rightarrow_d N(0, \frac{1+2\sum_{i=1}^q \rho_i^2}{T})$ para $k \geq q$. Outros resultados baseados em suposições mais gerais existem, por exemplo se r_t é um processo linear, veja

³De uma maneira informal, ergodicidade é a propriedade que nos permite estimar quantidades de interesse da série, como a média ou a autocorrelação, utilizando uma única trajetória do processo, isto é, a série de log-retornos observada. Veja definição no Capítulo 4.

em Box, G. E. P., Jenkins, G. M. e Reinsel, G. (1994), e Brockwell, P. J. e Davis, R. A. (1991). Para amostras pequenas $\hat{\rho}_k$ é um estimador viciado de ρ_k . Quando a distribuição F tem caudas pesadas (como é o caso de log-retornos) o erro de estimação pode ser até maior que as próprias estimativas (Davis, R. A. e Resnick, S. I. (1996) e Davis, R. A. e Mikosch, T. (1998)).

A função $\hat{\rho}_k$ para $k = 1, 2, \dots$ é a *função de autocorrelação amostral* (f.a.c. amostral) de r_t . Sendo $\rho_k = \rho_{-k}$, em geral examinamos apenas os coeficientes dos *lags* positivos. Como ρ_k mede apenas a força da dependência *linear* entre as variáveis envolvidas, a f.a.c. caracteriza a estrutura de dependência de processos lineares estacionários. Quando a cópula da distribuição conjunta dos retornos é a Gaussiana, a f.a.c. amostral caracteriza perfeitamente a estrutura de dependência de $\{r_t\}$. Fora do mundo Gaussiano, e/ou quando as séries são não estacionárias, apresentam sazonalidades, saltos, distorções, agregação, etc, esta função é menos importante, veja Guegan (2003).

A teoria de que o mercado não é previsível se traduz na *não* existência de correlações seriais. Contudo, a maneira como as ações são precificadas e os índices são compostos podem induzir algumas autocorrelações de curto prazo. Por isto precisamos saber modelar esta característica.

Na Figura 1.3 vemos as funções de autocorrelação amostral dos log-retornos diários dos índices do Brasil e do México. Nos gráficos retiramos o valor $\rho_0 = 1$ e incluímos o intervalo de 95% de confiança (linhas horizontais). A f.a.c. amostral do índice brasileiro indica que não deve haver autocorrelação para k pequeno, entretanto parece sugerir existência de memória longa para os retornos. Para o México, os dois primeiros coeficientes de autocorrelação são estatisticamente diferentes de zero ao nível de 5%, mas a maioria dos outros coeficientes está dentro dos limites de dois erros padrões.

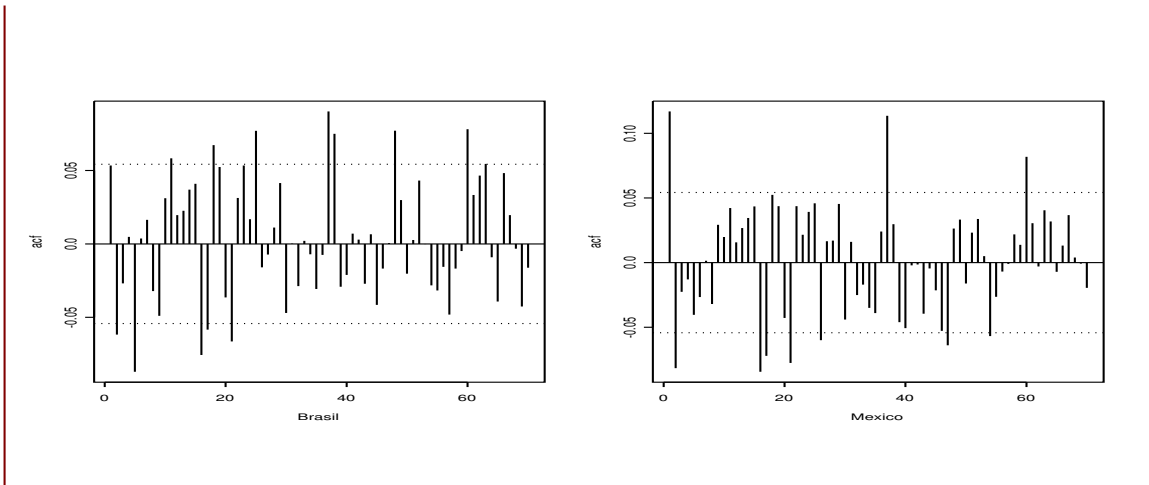


Figura 1.3: A a.c.f. amostral dos log-retornos dos índices Brasil e México.

Assim, a f.a.c. amostral nos dá uma idéia da extensão e do grau da memória do processo, sendo uma ferramenta importante para a identificação do processo gerador de uma série temporal, além de ser base para a construção de estatísticas teste. Em muitas aplicações em finanças se faz necessário testar se existe memória curta, isto é, se as m primeiras autocorrelações $\rho_1, \rho_2, \dots, \rho_m$ são zero. Box,

G.E.P. e Pierce, D.A. (1970) propuseram a estatística Portmanteau

$$Q^*(m) = T \sum_{k=1}^m \hat{\rho}_k^2$$

para testar

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_m = 0 \text{ vs } H_1 : \rho_i \neq 0 \text{ para algum } i \in \{1, \dots, m\}.$$

Sob a suposição de que $\{r_t\}$ é uma seqüência i.i.d., $Q^*(m) \rightarrow^d \chi_m^2$.

Ljung, G.M. e Box, G.E.P. (1978) modificaram a estatística $Q^*(m)$ aumentando o poder do teste em amostras finitas, propondo

$$Q(m) = T(T+2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{T-k}.$$

A escolha de m é muito importante, e em geral m é pequeno pois estamos testando se não existe memória curta. Estudos de simulação indicaram que $m \approx \log(T)$ é o valor que resulta em melhor performance do teste em termos de potência.

1.3 Fatos Estilizados de Séries de Retornos

O modelo básico em finanças assume que o preço de um ativo no tempo t , P_t , segue um movimento Browniano, solução de uma equação diferencial estocástica de Itô (veja detalhes no Apêndice 1.1 deste capítulo). Este modelo serve de base para vários outros, como por exemplo, o modelo de precificação de Black-Scholes. Uma consequência deste modelo é que os log-retornos seriam Normais i.i.d. Entretanto, as séries financeiras exibem outras características. O gráfico dos preços diários de fechamento na Figura 1.1 revela alguma semelhança (ver Nefcti (1996)) com a trajetória de um martingale contínuo com saltos (definição no Apêndice 1.1). Já para as séries de log-retornos, os gráficos no tempo (veja Figura 1.4) sugerem estacionariedade, apesar de apresentarem períodos de baixa e de alta volatilidade, com provável mudança de regime de volatilidade, não tão evidente talvez devido à presença de alguns valores extremos. De fato, em geral para séries de retornos aceitamos a hipótese nula de estacionariedade, mas os resultados de estatísticas e testes básicos (como os da Tabela 1.1), juntamente com os gráficos da f.a.c. amostral, quase sempre indicam *não* serem os log-retornos i.i.d. nem Normais.

Séries de retornos financeiros apresentam certas características comuns, chamadas de fatos estilizados. Comprovados empiricamente, eles sugerem que essas séries

1. São estacionárias.
2. Apresentam média amostral próxima de zero.
3. Apresentam distribuição (não condicional) aproximadamente simétrica.
4. Apresentam distribuição (não condicional) *não* Normal.

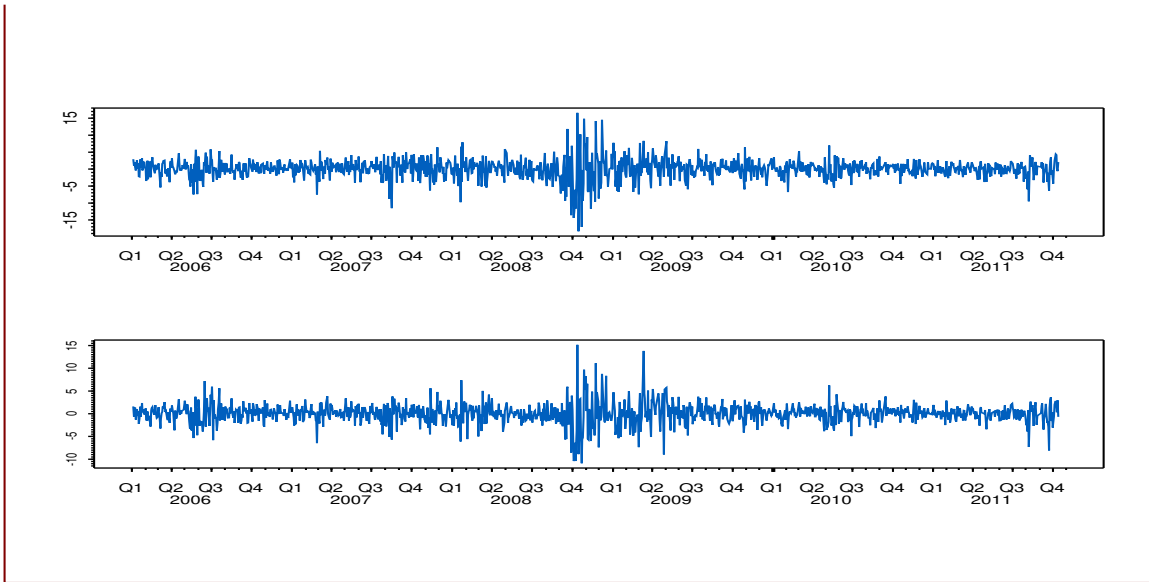


Figura 1.4: Gráfico no tempo dos log-retornos dos índices do Brasil e do México.

5. Apresentam excesso de curtose positivo (caudas pesadas).
6. Apresentam caudas com pesos diferentes (medidas, por exemplo, através do valor de ξ da distribuição de valores extremos, ou pelo índice de cauda α . Se $P(r > y)$ quando $y \rightarrow \infty$ é aproximadamente igual a $y^{-\alpha}$ com $\alpha \in (3, 5)$ fica caracterizada cauda pesada. Notemos, por exemplo, que se $\alpha \leq 3$ o terceiro momento não é finito (ver Capítulo 2).
7. Apresentam alguns pontos bem mais extremos que a maioria, os quais podem induzir assimetria para uma dada amostra.
8. Apresentam clusters de volatilidade (heteroscedasticidade condicional).
9. Apresentam alguma forma de não-linearidade (Pode surgir como consequência do fato de retornos financeiros reagirem diferentemente a choques positivos e negativos, ou a choques grandes ou pequenos, ou devido à dependência de cauda.).
10. Não apresentam autocorrelação significativa. Quando existe, é não zero e pequena apenas para os primeiros *lags*.
11. Apresentam autocorrelação nos retornos ao quadrado significativa para vários *lags*. São todas positivas e podem decair lentamente.
12. Apresentam memória longa na média e na volatilidade (decaimento lento da f.a.c. amostral (ver Capítulo 4)).

Por exemplo, para o índice do Brasil as figuras e estatísticas já analisadas confirmaram a presença dos fatos estilizados números 1, 4, 5, 7, 8, 10. O item 11 pode ser verificado efetuando-se o teste de Ljung-Box para os quadrados dos retornos. Para o índice brasileiro a hipótese nula de não autocorrelação até o *lag* 7 ou 10, foi rejeitada com p-valor igual a zero. Ferramentas para a análise dos fatos estilizados números 6, 9, e 12 serão fornecidas nos capítulos 2 e 4.

Outras características usualmente empiricamente confirmadas são: séries diárias de retornos de ações, índices, ou carteiras, raramente apresentam tendências ou sazonalidades, as quais podem aparecer em retornos intradiários; taxas de câmbio e taxas de juros podem apresentar tendências variando no tempo.

1.4 Correlações e Dependências entre Extremos

A f.a.c. amostral pode não ser a melhor ferramenta para descrever a dependência entre retornos financeiros, em particular entre seus extremos, já que captura apenas a dependência linear. Covariância e correlação são momentos da distribuição, mas a informação sobre os extremos se encontra nas caudas das distribuições. Um agravante já ressaltado é que não é acurada a estimação da autocorrelação para séries com caudas pesadas.

Para os modelos elípticos, a estrutura de dependência canônica é a matriz de covariâncias Σ , a qual não representa adequadamente a associação entre os extremos, a menos que estejamos no mundo Gaussiano, pois neste caso os quantis serão assintoticamente independentes. Contudo, mesmo fora do modelo Normal, a f.a.c. *ainda é útil* porque os modelos mais comuns, por exemplo ARIMA, são construídos de tal forma que a estrutura de segunda ordem do modelo é bem descrita pela f.a.c.

É sempre interessante calcular outras medidas de dependência entre variáveis, além do coeficiente de correlação linear. Seja um vetor aleatório bidimensional (r_i, r_j) com distribuição conjunta F_{ij} . O coeficiente de correlação baseado em postos de Kendall, $\tau_{ij} = \tau(r_i, r_j)$, é definido como

$$\tau(r_i, r_j) = P((r_i - r_i^*) \cdot (r_j - r_j^*) > 0) - P((r_i - r_i^*) \cdot (r_j - r_j^*) < 0), \quad (1.1)$$

onde (r_i, r_j) e (r_i^*, r_j^*) são dois vetores aleatórios independentes com distribuição F_{ij} . Este coeficiente pode ser entendido como uma medida da *concordância* entre dois vetores bivariados, já que (1.1) é a diferença entre as probabilidades de concordância e de discordância entre dois vetores aleatórios com uma mesma distribuição. É fácil ver que τ pode ser definido como

$$\tau(r_i, r_j) = E[\text{sign}((r_i - r_i^*)(r_j - r_j^*))]. \quad (1.2)$$

Para $d \geq 2$ a matriz de coeficientes τ define uma matriz de covariâncias positiva definida. O estimador empírico de τ pode ser calculado como

$$\hat{\tau}_{ij} = \left(\frac{T}{2} \right)^{-1} \sum_{1 \leq t < s \leq T} \text{sign}((r_{t,i} - r_{s,i})(r_{t,j} - r_{s,j})), \quad (1.3)$$

que é igual a

$$\hat{\tau}_{ij} = \left(\frac{T}{2} \right)^{-1} (C_T - D_T) \quad (1.4)$$

onde C_T e D_T representam o número de pares concordantes e discordantes na amostra de tamanho T .

O coeficiente de correlação baseado em postos de Spearman, $\rho_S(r_i, r_j)$, é definido como

$$\rho_S(r_i, r_j) = \rho(F_i(r_i), F_j(r_j)) , \quad (1.5)$$

onde F_i e F_j são as distribuições marginais de r_i e r_j . Quaisquer transformações monótonas crescentes r_i e de r_j levam ao mesmo conjunto de *rank*s ou postos. A Figura 1.5 ilustra esse fato e tem na parte superior o gráfico de 30 observações de uma Normal bivariada (X, Y) com médias zero, variâncias um e $\rho = 0.7$, e à direita o gráfico dos postos dessas observações. Na parte inferior temos o gráfico das transformações $\exp(2 * X)$ e $\exp(Y)$ e de seus postos. A figura também ilustra a influência das distribuições marginais na relação linear das variáveis.

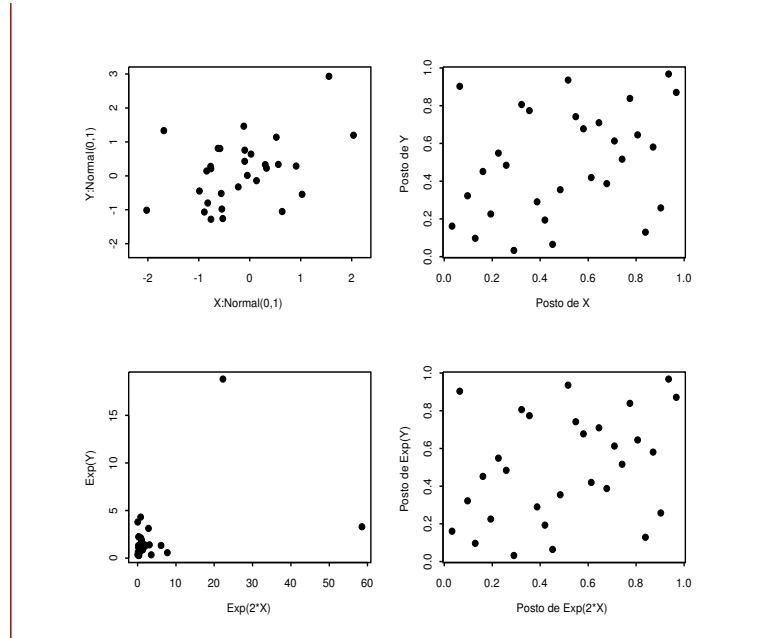


Figura 1.5: Na parte superior temos 30 observações da Normal bivariada (X, Y) com $\rho = 0.7$ e seus postos. Na parte inferior temos as transformações $\exp(2 * X)$ e $\exp(Y)$ e seus postos.

Assim como o coeficiente de correlação de Pearson ρ , o coeficiente de correlação de Spearman ρ_S é zero quando as variáveis são independentes. Contudo ρ_S apresenta várias vantagens em relação a ρ (veja Embrechts, P., McNeil, A. J. e Straumann, D. (2002)). Por exemplo, atinge os limites ± 1 se e somente se as variáveis são funcionalmente dependentes (equivale a dizer que a cópula que as liga é uma das cópulas limite de Fréchet – Hoeffding, ver Capítulo 3). Para o coeficiente de correlação linear os limites ± 1 podem ser atingidos quando uma variável é função linear da outra, uma restrição muito maior. Além disto ρ_S está sempre bem definido ao passo que ρ precisa que os segundos momentos das marginais sejam finitos.

Suponha uma amostra de tamanho T , $((r_{i,1}, r_{j,1}), \dots, (r_{i,T}, r_{j,T}))$ do vetor (r_i, r_j) . Substituímos os elementos de cada coluna por seus postos, números entre 1 e T , os quais denotaremos por R e S . Quanto mais próximos esses números, mais forte a estrutura de dependência entre as v.a.'s. Sejam $((R_1, S_1), \dots, (R_T, S_T))$ os postos para esta amostra, os quais estão bem definidos já que a probabilidade de ocorrência de empates é zero sob a suposição de continuidade para as v.a.'s. Dentre todas as funções que são invariantes sob transformações monótonas crescentes, os postos amostrais são as estatísticas que guardam a maior quantidade de informação sobre a estrutura de dependência dos dados (Oakes, D., 1982). Veja em Mendes, B. V. M. e Leal, R. P. C. (2010) uma investigação sobre a utilidade desta matriz de postos em finanças ao obter replicações semi-paramétricas da fronteira eficiente de Markowitz. Notemos que os postos podem ser equivalentemente dados pelos valores $(1/(T+1), \dots, T/(T+1))$. O estimador empírico de ρ_S é dado por

$$\rho_{S,T} = \frac{12}{T(T+1)(T-1)} \sum_{i=1}^T R_i S_i - 3 \frac{T+1}{T-1}. \quad (1.6)$$

Para medir a associação entre valores extremos de uma distribuição conjunta temos o coeficiente de dependência de cauda λ . Se existir associação positiva (Tawn, J. (1988a), Tawn, J. (1988b)) entre eventos extremos de r_1 e r_2 , então a probabilidade condicional $P(r_1 > F_1^{-1}(1-\alpha) | r_2 > F_2^{-1}(1-\alpha))$ é maior que zero e decresce quando $\alpha \downarrow 0$. A dependência de cauda superior λ_U entre r_1 e r_2 é definida como Joe, H. (1997)

$$\lim_{\alpha \rightarrow 0^+} \lambda_U(\alpha) = \lim_{\alpha \rightarrow 0^+} P(r_1 > F_1^{-1}(1-\alpha) | r_2 > F_2^{-1}(1-\alpha)),$$

se este limite existir. Para α pequeno, esta probabilidade não depende da ordem das duas variáveis r_1 e r_2 , as quais são assintoticamente dependentes na cauda superior se $\lambda_U \in (0, 1]$, e assintoticamente independentes se $\lambda_U = 0$.

Uma melhor interpretação deste conceito em finanças pode ser obtido se reescrevermos a definição de λ_U como

$$\lambda_U = \lim_{\alpha \rightarrow 0^+} P(r_1 > \text{VaR}_\alpha(r_1) | r_2 > \text{VaR}_\alpha(r_2)), \quad (1.7)$$

onde $\text{VaR}_\alpha(r_1) = F_1^{-1}(1-\alpha)$ é o Valor-em-Risco, medindo o risco nos ganhos conjuntos. Temos uma definição equivalente para perdas conjuntas extremas, levando ao conceito de coeficiente de dependência de cauda inferior, λ_L . Podemos interpretá-lo, por exemplo, como a probabilidade que um índice representativo de um mercado apresente uma queda mais extrema que o seu VaR_α , dado que um índice de outro mercado tenha ultrapassado o seu VaR_α , decresce quando α vai para zero, mas tem um limite, o qual pode ser maior que zero e é independente da ordem dos mercados.

1.5 Apêndice do Capítulo 1

AP1.1: Martingales. De uma maneira muito informal, um processo S_t é um martingale se os seus movimentos futuros, dado um conjunto de informações passadas I_t , são completamente não previsíveis. Veja definições formais em Nefcti (1996) ou Karatzas, I. e Shreve, S.E. (1988). Assim, de acordo com a definição, se S_t é um martingale, então a esperança da variação de S_t em um intervalo de tempo

$u > 0$, isto é, $E[S_{t+u} - S_t]$ deve ser zero. Isto significa que as direções dos movimentos futuros de um martingale são impossíveis de se prever. Martingales devem mostrar uma trajetória bastante irregular. Se um processo apresentar claramente tendências de curto ou longo prazo não deve ser um martingale.

Como em geral se espera que o preço de um ativo financeiro, P_t , cresça com o tempo, ele não é completamente não previsível, e portanto não deve ser um martingale. Seja Δ um intervalo de tempo pequeno e μ uma taxa de retorno positiva. Pode-se mostrar (Nefcti, 1996) que vale a aproximação: $E[P_{t+\Delta} - P_t] \cong \mu\Delta$. Assim, a maioria dos ativos financeiros não são martingales, mas podem ser “convertidos” em martingales, como exemplificamos a seguir.

Na ausência de possibilidades de arbitragem, o equilíbrio de mercado sugere que o preço de um ativo descontado pelo valor da taxa de um ativo livre de risco r_0 (por exemplo, o Certificado de Depósito Interbancário, CDI) durante um período de tempo u , se comporte como um martingale, já que $E[e^{-r_0 u} P_{t+u} | I_t] = P_t$, $u > 0$. Esta é a razão pela qual martingales têm um papel fundamental na precificação de ativos. É uma teoria muito rica, que inclui a possibilidade de estudo de fenômenos aleatórios em tempo contínuo, como o da Figura 1.1. Por exemplo, um processo P_t cujos incrementos ΔP_t sejam normalmente distribuídos (com média $\mu\Delta$ e variância $\sigma^2\Delta$) é um movimento Browniano. Subtraindo uma função determinística crescente no tempo μt pode-se transformar esse processo num martingale $P_t - \mu t$.

Entre os diversos tipos de martingales, provavelmente o mais simples deles seja o *passeio aleatório com incrementos i.i.d.* Neste caso, o processo log-preço p_t é dado pela seguinte equação

$$p_t = \mu + p_{t-1} + \epsilon_t$$

onde ϵ_t são i.i.d. com distribuição F , com média zero e variância σ^2 . Aqui μ é a variação de preço esperada (ou *drift*). As inovações, sendo independentes, garantem “um jogo honesto” (isto é, a esperança zero de um martingale), mas num sentido ainda mais forte do que o de um martingale. A suposição de incrementos i.i.d. implica em variações de preço com esperança constante e retornos não correlacionados, mas também que quaisquer funções não lineares dos retornos sejam também não correlacionadas (veja Campbell, J. Y., Lo, A. W. e MacKinley, A. C (1997)). Se F é a distribuição normal, então p_t é equivalente ao movimento Browniano amostrado em intervalos regulares.

1.6 Exercícios

Exercício 1: Obtenha os preços de fechamento diários de 4 ações ou índices de mercado para um período de pelo menos 5 anos. Calcule os log-retornos e os retornos simples das séries.

- (i) Para cada uma das séries de log-retornos, calcule a média e o desvio-padrão amostrais, o coeficiente de assimetria, o excesso de curtose, e uma estimativa para o coeficiente de autocorrelação de primeira ordem. Tome os quadrados dos retornos e para esta série calcule também uma estimativa para coeficiente de autocorrelação de primeira ordem. Compare.
- (ii) Divida as séries de log-retornos em 5 subperíodos de mesmo tamanho e calcule as mesmas estatísticas do item (i). Elas são estáveis ao longo do tempo? Aplique também testes estatísticos adequados.

- (iii) Repita o item (i) para as séries de retornos simples. O uso de uma definição diferente para retornos financeiros muda substancialmente as inferências feitas?
- (iv) Repita o item (ii) para as séries de retornos simples.
- (v) Identifique os fatos estilizados nas séries de log-retornos. Faça todas as análise gráficas e os testes estatísticos pertinentes.

Exercício 2: Para as mesmas séries do Exercício 1 e para os itens (ii) e (iv), calcule a “razão de variabilidade” para cada ano da amostra, com o objetivo de verificar a estabilidade da volatilidade ao longo do período amostrado.

Nota: Seja s uma estimativa de volatilidade, e $s\sqrt{264}$ a volatilidade anualizada. A razão de variabilidade no ano j é definida como a razão

$$\frac{\text{Volatilidade do periodo completo anualizada} / \text{Numero de anos no periodo}}{\text{Volatilidade do ano } j \text{ anualizada}}.$$

Exercício 3: Para as mesmas séries do Exercício 1 e para os mesmos quatro itens, obtenha estimativas para a locação e escala não condicionais usando o par de estimadores robustos: mediana amostral e MAD (*median of absolute deviations*). Compare os resultados com os obtidos usando os estimadores clássicos: média e desvio-padrão amostrais.

Nota: Para uma amostra (x_1, x_2, \dots, x_n) o MAD é definido como

$$MAD(x_1, x_2, \dots, x_n) = \text{mediana}\{|x_1 - \text{mediana}(x_j)|, \dots, |x_n - \text{mediana}(x_j)|\}$$

onde $\text{mediana}(x_j) = \text{mediana}\{x_1, x_2, \dots, x_n\}$ representa a mediana do conjunto de dados, e $|y|$ representa o valor absoluto de y .

Exercício 4: Usando os dados do Exercício 1 calcule o coeficiente de autocorrelação amostral de lag k , para $k = 1, 2, \dots, 12$, para as séries de log-retornos e para as séries dos quadrados dos log-retornos. Para cada ativo, faça o gráfico da f.a.c. amostral (até o lag 12) para as séries de retornos e para a série dos retornos ao quadrado, e verifique que há sempre mais correlação no segundo momento. Efetue o teste de Ljung-Box ($m = 12$) para uma constatação formal.

Exercício 5: Usando os dados do Exercício 1, investigue se algum dia da semana apresenta maior volatilidade dos que os outros. Para tanto, para cada ativo e para os dados de cada dia da semana, calcule os estimadores clássico e robusto de escala.

Exercício 6: Considere o índice Brasileiro de 03 de janeiro de 2006 a 13 de outubro de 2011. Responda as seguintes perguntas:

- (i) Qual é o retorno médio anualizado no período?
- (ii) Caso uma pessoa houvesse investido R\$100,00 no índice Brasil durante todo esse período, qual seria o seu valor no final, assumindo que não há custos de transação?

Exercício 7: Considere os retornos contínuos do índice Brasileiro de 03 de janeiro de 2006 a 13 de outubro de 2011. Use o nível de confiança de 5% e faça os testes a seguir.

(i) Teste a hipótese nula de que a medida de assimetria dos retornos é zero.

(ii) Teste a hipótese nula de que o excesso de curtose dos retornos é zero.

Exercício 8: Procure 6 artigos importantes na sua área de pesquisa, publicados em algum periódico classificado como A pela CAPES, e que tenha um enfoque empírico, ou que apresente uma análise de algum conjunto de dados. Retire deles uma lista das ferramentas utilizadas na análise exploratória feita, incluindo gráficos e testes, e verifique a utilidade delas na sua série de interesse.

Capítulo 2

Modelagem Não Condicional Univariada

Neste capítulo iremos abordar o problema da modelagem estatística não condicional de séries de retornos financeiros. A modelagem (multivariada) de um conjunto de d dessas séries financeiras será tema do Capítulo 3.

A forma da distribuição dos retornos de ativos financeiros é uma suposição crucial em vários modelos, como na teoria de Markowitz para carteiras eficientes, modelos teóricos para apreçamento de ativos, ou na modelagem de sinistros em Atuária. Por exemplo, o entendimento do comportamento da volatilidade é crucial nos modelos para apreçamento de opções. Lembremos que os testes estatísticos são efetuados sob a suposição de alguma distribuição. Sem dúvida, a suposição mais conveniente para a teoria e para análises empíricas de ativos é a da Normal multivariada com parâmetros que são constantes ao longo do tempo. Esta suposição é incentivada pelo fato de ser a distribuição Normal estável sob adição. Assim, qualquer carteira formada por ativos arbitrários também será normalmente distribuída.

Também no plano univariado a adequacidade da distribuição Normal como a distribuição subjacente (não condicional) de retornos financeiros tem sido questionada desde os trabalhos de Mandelbrot (1963) e Fama (1965). Muitas distribuições alternativas tem sido propostas na literatura para a modelagem da aleatoriedade observada em ativos ou carteiras, por exemplo a classe das distribuições estáveis e a classe bastante flexível das distribuições elípticas assimétricas que permitem discriminar o comportamento de perdas e ganhos com caudas mais pesadas que as da Normal. Outra opção seria a distribuição Weibull parcialmente assimétrica de Mittnik, S., Paoletta, M. e Rachev, S. (1998), onde são também discutidas as boas propriedades das distribuições estáveis.

Entretanto, mesmo utilizando distribuições com caudas mais pesadas para modelar retornos financeiros, sabe-se que as medidas de risco ficam subestimadas para probabilidades de ocorrência *muito pequenas*. Portanto, é necessário utilizar modelos que ajustem com precisão as caudas das distribuições, onde os dados são geralmente escassos. A Teoria de Valores Extremos (TVE) é capaz de fornecer uma forma paramétrica para as caudas de uma distribuição, permitindo assim extrapolações

para além dos limites dos dados. Focando nas caudas, os modelos da TVE não fazem suposições sobre os retornos *usuais*, aqueles situados no centro da distribuição. Devemos notar, porém, que qualquer substituição da premissa de normalidade deve manter um pouco de sua fácil implementação.

Considere uma série de retornos financeiros $\{r_1, r_2, \dots, r_T\}$ e sua f.d.a. indexada por um vetor de parâmetros $\theta \in \Theta$

$$F(r_1, r_2, \dots, r_T; \theta), \quad (2.1)$$

onde, por simplicidade, não levamos em consideração a possibilidade de também explicar o comportamento dos retornos a partir de outras variáveis representando grandezas da economia. O objetivo é especificar corretamente a forma de F e estimar θ , permitindo fazer previsões sobre valores futuros de r_t , simular trajetórias, e também entender seu comportamento passado.

A distribuição conjunta (2.1) pode ser escrita como

$$F(r_1, r_2, \dots, r_T; \theta) = F(r_1)F(r_2|r_1) \cdots F(r_T|r_{T-1}, \dots, r_1) = F(r_1) \prod_{t=2}^T F(r_t|r_{t-1}, \dots, r_1), \quad (2.2)$$

decomposição esta que enfatiza as dependências temporais de r_t . Se os retornos forem v.a.'s contínuas, então a densidade correspondente a (2.2) é

$$f(r_1, \dots, r_T; \theta) = f(r_1) \prod_{t=2}^T f(r_t|r_{t-1}, \dots, r_1). \quad (2.3)$$

Esta expressão mostra a importância da distribuição condicional na modelagem de séries temporais (Capítulo 4). Conforme antecipado no Capítulo 1, ela é usada para estimação por máxima verossimilhança.

Especificar corretamente como esta distribuição condicional $F(r_t|r_{t-1}, \dots)$ evolui no tempo será tópico do Capítulo 4. Obviamente, diferentes especificações levarão à diferentes modelagens. Por exemplo, a mais simples de todas seria uma das versões da hipótese de *passeio aleatório* (PA) (veja definição no Capítulo 4) que resulta em incrementos independentes. Se os log-preços, $p_t = \log(P_t)$ seguirem um processo estocástico tipo PA, então suas variações formariam uma série de v.a.'s i.i.d. Neste caso os retornos seriam independentes e a distribuição condicional seria igual a distribuição marginal:

$$F(r_t|r_{t-1}, \dots, r_1) = F(r_t),$$

quando então os retornos não seriam previsíveis. É esta distribuição marginal $F(r_t)$, ou simplesmente F (subjacente, não condicional, mas não necessariamente sob independência), que iremos investigar neste capítulo.

Na literatura existem várias propostas para a escolha de F , por exemplo sugerindo o uso da distribuição Normal, Log-Normal, distribuições estáveis, misturas de Normais (escalas diferentes), t -student, t -student assimétrica, etc. Uma revisão de distribuições assimétricas com caudas pesadas pode ser vista em Aas, K. e Haff, I. H. (Spring 2006), onde os autores propõem o uso de um caso especial da distribuição generalizada hiperbólica, único nesta classe capaz de modelar distintamente as caudas esquerda e direita, podendo uma ter decaimento polinomial e a outra um decaimento exponencial. Faremos a seguir uma revisão dos modelos mais adequados e/ou mais utilizados para representar a distribuição marginal dos retornos.

2.1 Distribuição Normal

A suposição feita é que os retornos simples R_t formem uma seqüência de v.a.'s i.i.d. Normais, com média (μ) e variância (σ^2) constantes, $N(\mu, \sigma^2)$. Esta suposição, além de não ser muitas vezes verificada empiricamente, é bastante simplificada já que, por exemplo, os R_t são limitados inferiormente por -1 . A estimação dos parâmetros desconhecidos pode ser feita através dos *estimadores de máxima verossimilhança sob normalidade*, \bar{r} e S^2 . A estimativa crucial é de μ , por exemplo como input no algoritmo *Média e Variância de Markowitz* para a alocação ótima de ativos. Devido à assimetria induzida por algum valor extremo, outros estimadores robustos podem dar melhor resultado. Por exemplo, o estimador robusto de locação *mediana amostral* ou algum tipo de M-estimador. Estimadores robustos para a escala são também as vezes preferidos, como por exemplo o MAD, definido no Exercício 3 do Capítulo 1. Veja detalhes em Yohai, V.J. (1987), Mendes, B. V. M. e Tyler, D. E. (1996), Mendes, B. V. M. e Tyler, D. E. (1998), Mendes, B. V. M. (1999), e Berrendero, J., Mendes, B. V. M. e Tyler, D. E. (2007).

A Tabela 2.1 fornece as estimativas dos estimadores de máxima verossimilhança (EMV) sob normalidade \bar{r} e S para μ e σ para os índices analisados. Para uma melhor comparação, incluímos na tabela os estimadores robustos mediana amostral (\tilde{r}) e MAD, além de M-estimadores de locação ($\hat{\mu}_R$) e escala ($\hat{\sigma}_R$). Para o cálculo dos M-estimadores são usados como estimadores iniciais um M-estimador com ponto de ruptura 0.5, e a solução final é um S-estimador obtido a partir de um algoritmo de re-amostragem (exaustiva) aleatória. Podemos ver que a assimetria à esquerda (Brasil e Indonésia) influenciou o valor do estimador clássico, e que o M-estimador deu uma resposta diferente dos demais no caso da locação do índice da China.

Tabela 2.1: *Estimativas Robustas de μ e σ para os índices Brasil, México, Indonésia e China.*

Índice	\bar{r}	S	\tilde{r}	MAD	$\hat{\mu}_R$	$\hat{\sigma}_R$
Brasil	0.0417	2.7954	0.1873	1.9486	0.1827	1.9618
México	0.0214	2.1405	0.1243	1.4583	0.0907	1.4977
Indonésia	0.0770	2.0517	0.1037	1.4023	0.1675	1.4191
China	0.0431	2.2539	0.0454	1.5681	0.1307	1.6000

Convenções utilizadas na tabela: \bar{r} : média amostral; S : desvio padrão amostral; \tilde{r} : mediana amostral; MAD: mediana dos desvios absolutos amostrais; $\hat{\mu}_R$: estimador robusto de locação; $\hat{\sigma}_R$: estimador robusto de escala.

Teste de aderência (*goodness of fit*). Seja F_0 a distribuição hipotetizada e F a verdadeira distribuição dos retornos r_t . A estatística de Kolmogorov para testar a hipótese nula

$$H_0 : F = F_0$$

é definida como

$$D_T = \sup_{-\infty < x < \infty} | \hat{F}_T(x) - F_0(x) | \quad (2.4)$$

onde \hat{F}_T representa a distribuição empírica dos dados, $\hat{F}_T(x) = (\sum_{t=1}^T \mathcal{I}_{(r_t \leq x)})/T$, $x \in \mathbb{R}$, onde $\mathcal{I}_{(A)}$ é a função indicadora do evento A , isto é, $\mathcal{I}_{(A)} = 1$ se A ocorrer e zero caso contrário.

Como $T\hat{F}_T(x)$ tem distribuição Binomial com probabilidade de sucesso igual a $F(x)$, temos que $\hat{F}_T(x)$ é um estimador não viciado de $F(x)$, com variância $F(x)(1 - F(x))/T$, além de ser consistente:

$$\hat{F}_T(x) \xrightarrow{p} F(x),$$

e assintoticamente Normal para cada x (a notação \xrightarrow{p} representa convergência em probabilidade). Pelo teorema de Glivenko-Cantelli (Chow, Y. S. e Teicher, H., 1988b) temos que para amostras grandes a função de distribuição empírica (f.d.e.) se assemelha à verdadeira distribuição dos dados, isto é, $\sup_{-\infty < x < \infty} |\hat{F}_T(x) - F(x)| \xrightarrow{q.c.} 0$ quando $T \rightarrow \infty$.

A distribuição de D_T sob a hipótese nula é a mesma para todas as distribuições F_0 contínuas. Este teste está implementado no S-Plus e no R para várias F_0 de interesse neste livro, como a Normal, Chiquadrado, Exponencial, Gama, Log-Normal, t , Uniforme, e Binomial. Para outras distribuições que veremos mais a frente, como por exemplo a t -assimétrica e as distribuições de valores extremos, a estatística (2.4) pode ser facilmente implementada. Formas fechadas para a distribuição de D_T sob H_0 e uma tabela para $T \leq 100$ podem ser vistas em Birnbaum, Z. W (1952). A distribuição assintótica é bem conhecida, e o valor crítico k_α para o nível de significância $\alpha = 0.05$ pode ser calculado como $1.358/(\sqrt{T} + 0.12 + 0.11/\sqrt{T})$ (veja Bickel, P.J. e Doksum, K.A. (2001)). Rejeitamos para $D_T > k_\alpha$.

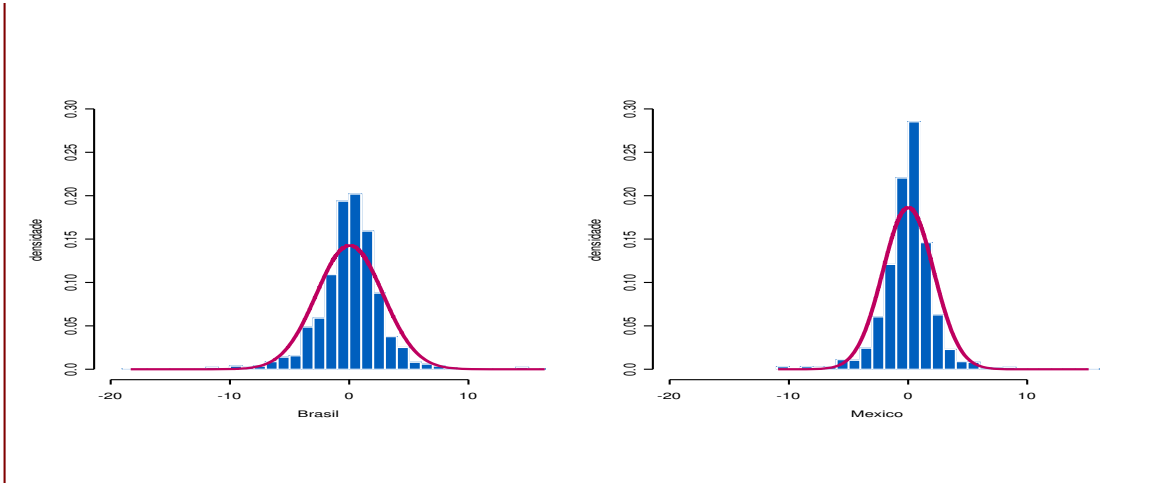


Figura 2.1: Histograma e densidade da Normal baseada nos estimadores de máxima verossimilhança para os índices do Brasil (esquerda) e do México (direita).

Conforme ilustra a Figura 2.1 a distribuição Normal parece não representar bem esses retornos financeiros. De fato, o teste de Kolmogorov dado acima rejeita fortemente a hipótese nula. O uso dos estimadores robustos para μ e σ não melhoram o ajuste.

2.2 Distribuição Log-Normal

Alternativamente podemos supor que os r_t são i.i.d $N(\mu, \sigma^2)$. Como $r_t = \log(1 + R_t)$, então $(1 + R_t)$ seria Log-Normal, e temos que

$$E[(1 + R_t)] = E[R_t] + 1 = \exp(\mu + \frac{\sigma^2}{2})$$

ou

$$E[R_t] = \exp(\mu + \frac{\sigma^2}{2}) - 1, \quad (2.5)$$

e

$$\text{var}(R_t) = \exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1). \quad (2.6)$$

Podemos expressar a esperança e a variância de r_t em função dos momentos $E[R_t]$ e $\text{var}(R_t)$ (Exercício 1 deste capítulo). Notemos que o retorno de k -períodos, $r_t[k]$, sendo soma de retornos de um período, também terá distribuição Normal. Conforme já vimos, para esta suposição não temos em geral evidências empíricas. Estimação pode ser feita através dos estimadores de máxima verossimilhança (EMV) sob normalidade utilizando-se a propriedade de invariância dos mesmos.

2.3 Distribuições Estáveis

Entre as distribuições mais propostas para modelar os retornos financeiros temos a família de distribuições estáveis. As distribuições estáveis também são chamadas de somas-estáveis devido ao fato de serem a classe de distribuições limite (não degeneradas) para somas padronizadas de variáveis aleatórias independentes e idênticamente distribuídas. São distribuições estáveis sob adição, uma propriedade requerida pelos retornos compostos continuamente r_t . Possuem uma representação espectral parametrizada pelo expoente característico α , $\alpha \in (0, 2]$, daí o nome de α -estáveis. Para $\alpha = 2$ temos a distribuição Normal, para $\alpha = 1$ temos a distribuição Cauchy. São capazes de modelar caudas pesadas mas possuem variância infinita (exceto a Normal).

Existem várias maneiras de se definir as distribuições estáveis. Uma v.a. (ou f.d., ou distribuição) é estável se dados números não negativos c_1 , c_2 e as v.a.'s i.i.d. X, X_1, X_2 , existem $b > 0$ e $a \in \mathbb{R}$ tais que

$$c_1 X_1 + c_2 X_2 \stackrel{d}{=} bX + a. \quad (2.7)$$

Se a é zero dizemos que X tem distribuição estritamente estável. A expressão (2.7) representa classes de distribuições que são fechadas (a menos de variações na locação e escala) sob convolução e multiplicação com números reais. Por exemplo, a convolução de duas distribuições Poisson é Poisson. Contudo, a distribuição de Poisson não satisfaz (2.7). Isso mostra que (2.7) é mais restritivo do que simplesmente ser uma classe fechada em relação à convolução.

Considere $S_n = X_1 + X_2 + \dots + X_n$ uma soma de v.a.'s i.i.d. com $X_i \sim G_i$ e seja $X \sim F$. Por (2.7) temos que se existirem constantes reais a_n e $b_n > 0$, $S_n = X_1 + \dots + X_n \stackrel{d}{=} b_n X + a_n$, podemos reescrever como

$$b_n^{-1}(S_n - a_n) \stackrel{d}{=} X \text{ quando } n \longrightarrow \infty,$$

e dizemos que as G_i 's pertencem ao domínio de atração da F . Novamente, se uma distribuição é estável, então ela é distribuição-limite para soma de v.a.'s i.i.d. De fato, o teorema a seguir mostra que não existem outras possibilidades para a distribuição-limite.

Teorema: (Propriedade limite das distribuições estáveis) A classe das distribuições estáveis (não degeneradas) coincide com a classe de todas as possíveis distribuições-limite (não degeneradas) para as somas de v.a.'s i.i.d. (apropriadamente centradas e normalizadas).

Existe a representação espectral de uma distribuição estável. Uma distribuição estável tem função característica para $t \in R$,

$$\Phi_x(t) = E[e^{itx}] = \exp\{-c|t|^\alpha(1 - i\beta \text{sign}(t)z(t, \alpha))\}$$

$$\text{onde } c > 0, \alpha \in (0, 2], \beta \in [-1, 1] \text{ e } z(t, \alpha) = \begin{cases} \ln(\frac{\pi\alpha}{|t|}), & \text{se } \alpha \neq 1 \\ -\frac{2}{\pi}\ln|t|, & \text{se } \alpha = 1. \end{cases}$$

Na representação acima a locação é zero. O caso $c = 0$ corresponde a uma distribuição degenerada. O parâmetro α determina as propriedades básicas desta classe de distribuição (como momentos, caudas, comportamento assintótico de somas, etc), e na função característica é chamado de expoente característico. A distribuição correspondente é chamada de α -estável.

Se $\alpha=2$ temos a distribuição Normal, e vale que $\Phi_x(t) = \exp\{-c^2 t^2\}$, a função característica de uma $N(0, 2c^2)$. Se $\alpha < 2$ pode ser assimétrica. Se $\alpha = 1$ temos a distribuição de Cauchy, a qual é importante em finanças. Para α fixo, parâmetro c está relacionado com variância. Para α fixo, β descreve a assimetria da distribuição. Notemos que $\Phi_x(t)$ é de valor real apenas se $\beta = 0$ (a v.a. X é simétrica).

Na literatura de finanças encontramos várias propostas do uso das distribuições estáveis. Em Fielitz, B. D. e Rozelle, J. P. (1983) os autores as comparam à misturas de Normais e desenvolvem testes baseados na propriedade de estabilidade sob adição.

2.4 Mistura de Normais

Com o objetivo de ao mesmo tempo manter a simplicidade matemática da distribuição Normal e modelar as caudas pesadas, alguns autores têm utilizado misturas de Normais com diferentes escalas. Neste caso a distribuição F seria dada por

$$(1 - \alpha)N(\mu, \sigma_1^2) + (\alpha)N(\mu, \sigma_2^2)$$

com $0 \leq \alpha \leq 1$. Colocando-se σ_1^2 pequeno e σ_2^2 grande, para α pequeno (digamos 0.05), estamos garantindo que uma grande proporção dos retornos tenham um comportamento *usual* e apenas uma pequena parcela poderia ser de grandes variações.

Evidências empíricas da validade deste modelo de misturas envolvendo de 1 a 5 distribuições Normais são fornecidas em Kon, S. J. (1984), e comparadas através de testes da razão de verossimilhanças (ver Apêndice [AP2.9] para definição de teste da razão de verossimilhanças). Em Buckley, I., Saunders, D. e Seco, L. (2008) os autores consideram o problema de otimização de uma carteira sob a suposição que a distribuição subjacente dos retornos é uma mistura de duas Normais multivariadas,

que representam diferentes estados da economia. Os autores obtêm uma caracterização geométrica da carteira ótima que reduz substancialmente a complexidade do problema de otimização.

2.5 Distribuição Normal Assimétrica

Existem várias maneiras de se construir a densidade de uma v.a. Normal Assimétrica. A grande motivação para a definição deste tipo de v.a. é obter uma classe de distribuições paramétricas que apresentem uma transição contínua da normalidade para a não-normalidade através de um parâmetro. Uma dessas propostas pode ser vista em Azzalini, A. (1985) e Azzalini, A. (1986), onde a densidade $\phi(\cdot, \cdot)$ da Normal Assimétrica é definida como

$$\phi(r; \lambda) = 2\phi(r)\Phi(\lambda r) \quad (2.8)$$

para $-\infty < r < \infty$, onde λ é o parâmetro que controla a assimetria, e onde $\phi(\cdot)$ e $\Phi(\cdot)$ representam respectivamente a densidade e a f.d.a. de uma Normal padrão, $N(0, 1)$. Notemos que se $\lambda = 0$, $\phi(r; \lambda)$ é a densidade de uma $Z \sim N(0, 1)$. Se $\lambda \rightarrow \infty$ temos o caso limite que corresponde à densidade da v.a. $|Z|$ (Half-Normal). A Figura 2.2 mostra a densidade (2.8) para $\lambda = 0$ em preto, $\lambda = \pm 1$ em azul, e $\lambda = \pm 4$ em rosa.

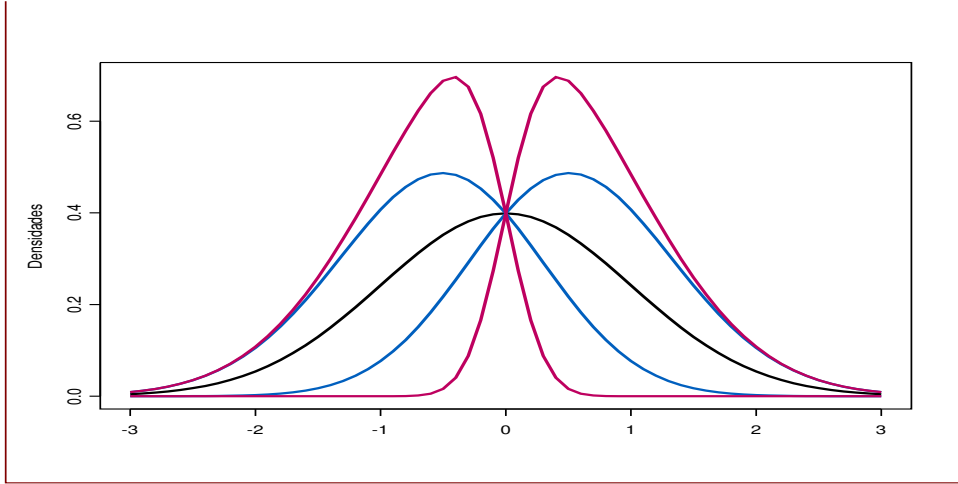


Figura 2.2: Densidade da Normal assimétrica para $\lambda = 0$ em preto, $\lambda = \pm 1$ em azul, e $\lambda = \pm 4$ em rosa.

A função geradora de momentos de uma v.a. r é dada por $M_r(s) = E[\exp(sr)]$, para $-s_0 \leq s \leq s_0$. Se $E[\exp(s_0 r)] < \infty$ para algum $s_0 > 0$, então $M_r(s)$ está bem definida. Estando $M_r(s)$ bem definida numa vizinhança $\{s : |s| \leq s_0\}$ de zero, todos os momentos de r são finitos e a função geradora de momentos tem derivadas de todas as ordens em $s = 0$, e os momentos $E[r^k]$, para $k = 1, 2, \dots$, são iguais às derivadas $\frac{d^k}{ds^k} M_r(s)$ aplicadas em $s = 0$.

A função geradora de momentos da Normal Assimétrica (2.8) é dada por

$$M(s) = 2 \exp(s^2/2) \Phi(\delta s),$$

onde $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$. Segue que o valor esperado da v.a. definida em (2.8) é dado por $\sqrt{\frac{2\delta}{\pi}}$ e a variância é dada por $1 - \frac{2\delta^2}{\pi}$. O coeficiente de assimetria é dado por $\frac{\sqrt{2}(4-\pi)\delta^3}{(\pi-2\delta^2)^{3/2}}$ e assume valores entre -1 e 1 . O excesso de curtose é dado por $\frac{8\delta^4(\pi-3)}{(\pi-2\delta^2)^2}$. Em aplicações é útil notar que se $X \sim \phi(x; \lambda)$ então $-X \sim \phi(x; -\lambda)$. Como se pode ver em Arellano-Valle, R. B., Gomez, H. W. e Quintana, F. A. (2003) o coeficiente de assimetria varia no intervalo $(-0.995, 0.995)$ e o valor máximo do excesso de curtose é 0.869 .

Os valores de excesso de curtose de uma Normal Assimétrica como definida acima não são suficientes para capturar o peso da cauda observado para a grande maioria das séries financeiras. Eles podem ser maiores se introduzirmos um parâmetro de escala. Assim, podemos definir uma v.a. Normal Assimétrica com mais dois parâmetros de locação e escala, μ e σ , a qual tem densidade

$$f(r; \mu, \sigma, \lambda) = \frac{2}{\sigma} \phi\left(\frac{r - \mu}{\sigma}\right) \Phi\left(\frac{\lambda(r - \mu)}{\sigma}\right).$$

A esperança e variância da v.a. definida pela densidade acima são dados por $\sqrt{\frac{2}{\pi}}\delta\sigma + \mu$ e $(1 - \frac{2}{\pi})\delta^2\sigma^2$. Estimação pode ser através do método da máxima verossimilhança. Outras referências importantes são Branco, M.D. e Dey, D.K. (2001), Azzalini, A. e Capitanio, A. (1999), Arellano-Valle, R. B. , Branco, M. D. e Genton, M. G. (2006).

2.6 Distribuições Elípticas

A distribuição empírica de retornos financeiros apresenta algumas similaridades com a distribuição Normal, por exemplo a forma de sino. Assim, uma possibilidade para F seria algum membro da família de distribuições elípticas, por ser esta uma família paramétrica flexível, unimodal, que generaliza a distribuição Normal, mantendo ainda algumas de suas importantes propriedades.

Daremos a definição na sua forma geral, para $d \geq 1$, pois basta fazer $d = 1$ para obter o caso univariado tratado neste capítulo. Começaremos com o caso particular das distribuições esféricas, já que a maioria das propriedades dessa família se estendem para o caso elíptico.

Distribuições esféricas. A família de *distribuições esféricas* são uma grande classe de distribuições de vetores aleatórios com *marginais simétricas idênticas não correlacionadas*. Um caso *particular* desta classe é a distribuição Normal d -variada $\mathbf{Z} \sim N_d(\mathbf{0}, \mathbf{I}_d)$, centrada em $\boldsymbol{\mu} = \mathbf{0}$ e onde \mathbf{I}_d é uma matriz diagonal unitária d -dimensional, o *único* membro desta classe que possui marginais *independentes*.

Um vetor aleatório tem distribuição esférica d -variada se para toda transformação ortogonal U (U é tal que $UU' = U'U = \mathbf{I}_d$) temos

$$U\mathbf{r} \stackrel{d}{=} \mathbf{r},$$

onde $\stackrel{d}{=}$ significa “tem a mesma distribuição de”. Isto é, a distribuição desses vetores é fechada quanto a rotações.

Se o vetor aleatório \mathbf{r} tem distribuição esférica então sua função característica, definida como $\phi_{\mathbf{r}}^*(\mathbf{t}) = E[\exp(it'\mathbf{r})]$, é dada por $\phi_{\mathbf{r}}^*(\mathbf{t}) = \psi(t'\mathbf{t}) = \psi(t_1^2 + \dots + t_d^2)$, para alguma função ψ de argumento escalar não negativo. ψ é chamado de *gerador característico* da distribuição esférica, e escrevemos

$\mathbf{r} \sim \mathcal{S}_d(\psi)$. Exemplo: A distribuição Normal multivariada padrão, $N_d(\mathbf{0}, \mathbf{I}_d)$ é esférica. Sua função característica é dada por $\phi^*(\mathbf{t}) = \exp(-\frac{1}{2}\mathbf{t}'\mathbf{t})$ e assim seu gerador característico é $\psi(t) = \exp(-\frac{1}{2}t)$.

Combinações lineares de vetores esféricos, a menos de mudanças na locação e escala, ainda pertencem à classe de distribuições esféricas, isto é, são do *mesmo tipo*. Esta propriedade é importante para mostrar a sub-aditividade do Valor-em-Risco quando os ativos componentes de uma carteira são provenientes de uma distribuição elíptica.

De modo equivalente, dizemos que a v.a. \mathbf{r} tem distribuição esférica se e somente se possui a representação estocástica

$$\mathbf{r} \stackrel{d}{=} R\mathbf{S}, \quad (2.9)$$

onde \mathbf{S} é distribuída uniformemente na esfera unitária $\mathcal{S}^{d-1} = \{\mathbf{s} \in \mathbb{R}^d : \mathbf{s}'\mathbf{s} = 1\}$ e $R \geq 0$ é uma v.a. radial independente de \mathbf{S} . Exemplo: Se $\mathbf{r} \sim N_d(\mathbf{0}, \mathbf{I}_d)$, então $\mathbf{r}'\mathbf{r} \sim \chi_d^2$, $R^2 \sim \chi_d^2$, e \mathbf{S} tem densidade na esfera unitária com $E[\mathbf{S}] = \mathbf{0}$ e $cov(\mathbf{S}) = \frac{\mathbf{I}_d}{d}$.

A densidade f de uma distribuição esférica deve ser da forma

$$f(\mathbf{r}) = g(\mathbf{r}'\mathbf{r}) = g(r_1^2 + \cdots + r_d^2) \quad (2.10)$$

para alguma função $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$, chamada de *gerador da densidade*. Assim, caso exista, a densidade é constante nas hipersferas centradas na origem, $\{\mathbf{r} : r_1^2 + \cdots + r_d^2 = c\} \in \mathbb{R}^d$.

Por exemplo, consideremos a distribuição normal padrão bivariada com $\rho = 0$. Todas as curvas de nível (curvas de densidade constante) desta distribuição serão circunferências. No caso de uma distribuição normal tri-variada padrão com marginais não correlacionadas, todas as suas “curvas de nível” teriam a forma de esferas. Assim, podemos afirmar que as distribuições esféricas são aquelas cujas densidades são constantes em esferas.

Além da normal, existem outras distribuições que pertencem a esta classe. Por exemplo, temos a distribuição t-student em \mathbb{R}^d com ν graus de liberdade. Notemos que aqui as marginais são v.a.s não-correlacionadas, mas não são independentes.

Vemos então que toda distribuição esférica pode ser gerada a partir de uma parte comum a elas, que é uma distribuição uniforme padrão no \mathbb{R}^d combinada com uma variável aleatória unidimensional que determinaria a distribuição resultante da mistura dessas duas. Por exemplo, no caso da distribuição normal, a variável geradora R é uma $\sqrt{\chi_d^2}$. Outro exemplo é dado pela distribuição t-student com ν graus de liberdade que tem sua variável geradora R dada por $R^2/d \sim \mathcal{F}(d, \nu)$, onde \mathcal{F} representa a distribuição F de Snedecor com graus de liberdade d e ν , inteiros positivos.

Distribuições elípticas. As distribuições elípticas aparecem como extensões das distribuições esféricas. Uma forma de se obter estas distribuições é introduzindo uma estrutura de correlação nas distribuições esféricas. Podemos definir da seguinte forma: Dizemos que \mathbf{r} tem distribuição *elíptica simétrica* (ou simplesmente distribuição elíptica) com vetor de locação $\boldsymbol{\mu}$ e matriz $d \times d$ de dispersão $\boldsymbol{\Sigma}$ se

$$\mathbf{r} \stackrel{d}{=} \boldsymbol{\mu} + A\mathbf{Y} \quad (2.11)$$

onde \mathbf{Y} tem distribuição esférica, $\mathbf{Y} \sim \mathcal{S}_k(\psi)$, $\boldsymbol{\mu} \in \mathbb{R}^d$ é um vetor de constantes, e A é matriz $d \times k$, sendo $\boldsymbol{\Sigma} = AA'$. Vemos que as distribuições elípticas podem ser obtidas através de transformações

afins de distribuições esféricas. Escrevemos $\mathbf{r} \sim \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$, onde ψ é o gerador característico da distribuição. Portanto, podemos escrever a função característica de uma distribuição elíptica como:

$$\begin{aligned}\psi(\mathbf{t}) &= E[\exp(i\mathbf{t}'\mathbf{r})] = E[\exp(i\mathbf{t}'(A\mathbf{Y} + \boldsymbol{\mu}))] = \\ &= \exp(i\mathbf{t}'\boldsymbol{\mu}) \exp(i(A'\mathbf{t})'\mathbf{Y}) = \exp(i\mathbf{t}'\boldsymbol{\mu}) \psi_{\mathbf{Y}}(\mathbf{t}'A) = \\ &= \exp(i\mathbf{t}'\boldsymbol{\mu}) \phi_{\mathbf{Y}}(\mathbf{t}'AA'\mathbf{t}) = \\ &= \exp(i\mathbf{t}'\boldsymbol{\mu}) \phi_{\mathbf{Y}}(\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}),\end{aligned}$$

Exemplo: a distribuição $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ pode ser escrita como $\mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi(\cdot))$ ou $\mathcal{E}_d(\boldsymbol{\mu}, c\boldsymbol{\Sigma}, \psi(\frac{\cdot}{c}))$ para $\psi(a) = \exp(-\frac{1}{2}a^2)$ e para algum $c > 0$ ($\boldsymbol{\Sigma}$ e ψ são definidos a menos de uma constante positiva).

Notemos que utilizando (2.9) podemos reescrever (2.11) como

$$\mathbf{r} \stackrel{d}{=} \boldsymbol{\mu} + R\mathbf{A}\mathbf{S}$$

onde \mathbf{S} é distribuída uniformemente na esfera unitária $\mathcal{S}^{d-1} = \{\mathbf{s} \in \mathbb{R}^d : \mathbf{s}'\mathbf{s} = 1\}$, $R \geq 0$ é uma v.a. radial independente de \mathbf{S} , e A é matriz $d \times k$ com $AA' = \boldsymbol{\Sigma}$, sendo $\boldsymbol{\Sigma}$ positiva definida. A relação entre as distribuições esféricas e elípticas é

$$\mathbf{r} \sim \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi) \iff \boldsymbol{\Sigma}^{-1/2}(\mathbf{r} - \boldsymbol{\mu}) \sim \mathcal{S}_d(\psi)$$

Assim, se \mathbf{Y} tem distribuição esférica, $\mathbf{Y} \sim \mathcal{S}_k(\psi)$, com gerador de densidade g , então $\mathbf{r} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{Y}$ tem densidade

$$f(\mathbf{r}) = \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} g((\mathbf{r} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{r} - \boldsymbol{\mu})).$$

É fácil ver que o valor da densidade de uma distribuição elíptica é constante sobre elipsóides em \mathbb{R}^d , os conjuntos $(\mathbf{r} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{r} - \boldsymbol{\mu}) = c$ para uma constante $c > 0$.

As distribuições Normais multivariadas, sendo caso particular das elípticas com vetor de médias $\boldsymbol{\mu}$, matriz de dispersão $\boldsymbol{\Sigma}$ e gerador ψ , exibem várias das propriedades das distribuições elípticas. Para detalhes veja Fang, K., Kotz, S. e Ng, W. (1990). Ressaltamos as seguintes propriedades: (i) transformações afins de um vetor elíptico $\mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$ têm distribuição elíptica com o mesmo gerador ψ ; (ii) as distribuições marginais de um vetor elíptico $\mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$ têm distribuição elíptica com o mesmo gerador ψ . Em particular as *marginais univariadas* F serão todas iguais; (iii) as formas quadráticas $(\mathbf{r} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{r} - \boldsymbol{\mu})$ na maior parte dos casos de interesse possuem distribuição conhecida; (iv) as distribuições condicionais ainda são elípticas; (v) a convolução de dois vetores elípticos com mesma matriz de dispersão $\boldsymbol{\Sigma}$ tem distribuição elíptica. É importante ter em mente todas essas propriedades ao se supor uma distribuição elíptica para modelar um conjunto de d ativos ou calcular medidas de risco.

O coeficiente de correlação é uma medida natural de dependência no universo das distribuições esféricas e elípticas. Isto ocorre devido aos seguintes fatos:

1. Qualquer combinação linear, ou transformação afim, de um vetor aleatório com distribuição elíptica tem também distribuição elíptica, e mais, com o mesmo gerador característico. Assim, se $\mathbf{r} \sim \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ e $B \in \mathbb{R}^{m \times d}$, $b \in \mathbb{R}^m$, então

$$B\mathbf{r} + b \sim \mathcal{E}_m(B\boldsymbol{\mu} + b, B\Sigma B^t, \phi).$$

2. As distribuições marginais¹ de vetores aleatórios elípticos são também elípticas com o mesmo gerador. Seja $\mathbf{r} = \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \end{pmatrix} \sim \mathcal{E}_n(\Sigma, \boldsymbol{\mu}, \phi)$ com $\mathbf{r}_1 \in \mathbb{R}^p, \mathbf{r}_2 \in \mathbb{R}^q, p + q = d$. Seja $E[\mathbf{r}] = \boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$, $\boldsymbol{\mu}_1 \in \mathbb{R}^p, \boldsymbol{\mu}_2 \in \mathbb{R}^q$ e $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$. Então,

$$\mathbf{r}_1 \sim \mathcal{E}_p(\boldsymbol{\mu}_1, \Sigma_{11}, \phi), \mathbf{r}_2 \sim \mathcal{E}_q(\boldsymbol{\mu}_2, \Sigma_{22}, \phi).$$

3. Assumindo que Σ seja estritamente positiva definida, temos que a distribuição condicional de $r_1|r_2$ é também elíptica, e que o melhor preditor linear para $r_1|r_2$ é o dado por uma regressão linear. Especialmente no caso da normal, temos que $\Sigma_{r_1|r_2}$ não depende de r_2 . Este último fato é de grande importância no ambiente da análise de regressão.

Como as distribuições marginais são do mesmo tipo que as conjuntas, podemos concluir que uma distribuição elíptica é unicamente determinada pelo conhecimento de sua média, matriz de covariância e gerador característico. O que implica que a estrutura de dependência, em outras palavras, a cópula pertinente a distribuições deste tipo é unicamente determinada pela matriz de correlação. Este fato é importante para a estimação e simulação de cópulas elípticas.

Assumimos o segundo momento finito e uma amostra aleatória $T \times d$ do vetor de retornos $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_d$ para estimar o vetor de médias $\boldsymbol{\mu}$, a matriz de dispersão Σ , e a matriz de correlações \mathbf{P} . Os estimadores amostrais $(\bar{\mathbf{r}}, S^2)$ já definidos podem ser estendidos para o caso $d > 1$, isto é, definimos $\bar{\mathbf{r}} = (\bar{r}_1, \dots, \bar{r}_d)$ e $\mathbf{S}^2 = 1/T \sum_{i=1}^T (\mathbf{r}_{i,\cdot} - \bar{\mathbf{r}})(\mathbf{r}_{i,\cdot} - \bar{\mathbf{r}})'$, onde $\mathbf{r}_{i,\cdot}$ representa a i -ésima linha da matriz de dados $T \times d$.

Temos que $\bar{\mathbf{r}}$ e \mathbf{S}^2 são estimadores não viciados e consistentes (sob suposições fracas, mesmo substituindo independência por “não correlacionados”) de $\boldsymbol{\mu}$ e da matriz de covariâncias. Contudo, podem não ser os melhores estimadores principalmente se as distribuições tiverem caudas pesadas ou existirem contaminações, quando então o vício para amostras finitas pode ser muito grande. Para um excelente estudo sobre as alternativas robustas eficientes com alto ponto de ruptura e função de influência limitada, veja Lopuhaä, H. P. (1991) e Lopuhaä, H. P. (1992).

M-estimadores robustos não viciados e eficientes, baseados em alguma função de peso redescendente (Mendes, B. V. M. e Tyler, D. E., 1996), são em geral obtidos a partir de estimadores iniciais com alto ponto de ruptura (próximo de 0.5) como inputs em algum algoritmo que procura melhorar a eficiência dos iniciais. Para mais detalhes veja Lopuhaä, H.P. e Rousseeuw, P.J. (1991), Lopuhaä, H. P. (1999), Berrendero, J., Mendes, B. V. M. e Tyler, D. E. (2007), Roelant, E., Van Aelst, S. e Croux, C. (2009).

Neste tipo de abordagem em dois passos o objetivo é identificar primeiramente os pontos discordantes do padrão de dependência definido pela maioria. Uma solução simples é identificar esses pontos através da distância de Mahalanobis e atribuir aos pontos com distância robusta maior que um certo

¹Isto é, qualquer subconjunto de dimensão k , $k \geq 1$, das marginais r_j , $j = 1, \dots, d$, subconjunto este denotado por $\mathbf{r}^k = (r_i, \dots, r_j), k < d$.

ponto de corte o peso zero, calculando no passo 2 a matriz de covariâncias clássica \mathbf{S}^2 utilizando apenas os pontos com peso 1. Em geral o ponto de corte é dado pelo quantil de probabilidade acumulada igual a 0.975 de uma v.a. qui-quadrado com 2 graus de liberdade. Por exemplo, uma opção para esta estimação robusta em dois passos, presente tanto no S-Plus como no R, usa (Rousseeuw, P. J., 1985) como estimador inicial o *Minimum Covariance Distance* (MCD, ver algoritmo em Rousseeuw, P. J. e Van Driessen, K. (1999)). Outra opção para o estimador inicial, também presente em ambos os softwares citados, é o estimador baseado em projeções de Stahel-Donoho (SD) (veja Stahel, W. A. (1981) e Donoho (1982)).

A tabela 2.2 compara as estimativas das locações e covariâncias, μ_i e γ_{ij} , $i, j = 1, 2, 3, 4$, para os índices Brasil (1), Mexico (2), Indonésia (3), China (4), e para os métodos (i) clássico, $\bar{\mathbf{r}}$ e \mathbf{S}^2 ; (ii) MCD com ponto de ruptura 0.5; (iii) em 2 passos tendo como estimador inicial o MCD; (iv) em 2 passos tendo como estimador inicial o Stahel-Donoho (SD). Observamos que as estimativas robustas em geral são próximas e bem diferentes das clássicas.

Tabela 2.2: *Estimativas das locações e covariâncias, μ_i e γ_{ij} , $i, j = 1, 2, 3, 4$, para os índices Brasil (1), Mexico (2), Indonésia (3), China (4), e para os métodos (i) clássico, $\bar{\mathbf{r}}$ e \mathbf{S}^2 ; (ii) MCD com ponto de ruptura 0.5; (iii) em 2 passos tendo como estimador inicial o MCD; (iv) em 2 passos tendo como estimador inicial o Stahel-Donoho (SD).*

Parâmetro	Método (i) Clássico	Método (ii) MCD	Método (iii) 2 passos - MCD	Método (iv) 2 passos SD
γ_{11}	7.8140	3.4811	3.0638	3.7739
γ_{22}	4.5816	2.1321	1.8587	2.3323
γ_{33}	4.2096	2.1039	1.7621	2.0637
γ_{44}	5.0800	2.2827	2.0265	2.4766
γ_{12}	4.9088	2.0106	1.8237	2.6446
γ_{13}	2.0676	0.2575	0.5329	0.6733
γ_{14}	2.9223	0.6156	0.7235	0.9363
γ_{23}	1.5604	0.3595	0.4255	0.6086
γ_{24}	2.1476	0.6478	0.5658	0.8000
γ_{34}	2.9513	1.1073	0.9908	1.2420
μ_1	0.0417	0.2413	0.1891	0.2067
μ_2	0.0214	0.1462	0.1004	0.1284
μ_3	0.0770	0.3108	0.1967	0.1976
μ_4	0.0431	0.2783	0.1459	0.1388

Outra maneira de estimar \mathbf{P} é estimar os coeficientes de correlação linear $\rho_{ij} = \rho(\mathbf{r}_i, \mathbf{r}_j)$ através dos coeficientes de correlação monótona Kendall, τ_{ij} , baseados nos *ranks* das observações e definidos no Capítulo 1.

Para distribuições elípticas vale que

$$\tau_{ij} = \frac{2}{\pi} \arcsin(\rho_{ij}). \quad (2.12)$$

A prova deste resultado se baseia na propriedade de fechamento das distribuições elípticas em relação às convoluções. Este resultado será utilizado no Capítulo 4 para a estimação de cópulas elípticas.

2.7 Distribuição *t*-Student

Um membro importante da família de distribuições elípticas é a distribuição *t-student* (ou simplesmente *t*). Novamente daremos a definição no caso multivariado, já que sua particularização para o caso $d = 1$ é direta. Dizemos que um vetor aleatório d -dimensional $\mathbf{r} = (r_1, r_2, \dots, r_d)$ tem distribuição *t*-multivariada com ν graus de liberdade, com vetor de médias $\boldsymbol{\mu}$ e matriz (positiva definida) de dispersão $\boldsymbol{\Sigma}$, se sua densidade é dada por

$$f(\mathbf{r}) = \frac{\Gamma(\frac{\nu+d}{2})}{\Gamma(\frac{\nu}{2})\sqrt{(\pi\nu)^d|\boldsymbol{\Sigma}|}} \left(1 + \frac{(\mathbf{r} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{r} - \boldsymbol{\mu})}{\nu}\right)^{-\frac{\nu+d}{2}} \quad (2.13)$$

e denotaremos por $\mathbf{r} \sim t_d(\nu, \boldsymbol{\mu}, \boldsymbol{\Sigma})$. É importante lembrar que a matriz de covariâncias de \mathbf{r} *não* é igual a $\boldsymbol{\Sigma}$, mas $\text{cov}(\mathbf{r}) = \frac{\nu}{\nu-2}\boldsymbol{\Sigma}$, a qual somente está definida para $\nu > 2$ (veja Kotz, S., Balakrishnan, N. e Johnson, N. (2000)).

A distribuição *t*-multivariada também pertence à classe de misturas de distribuições Normais multivariadas com diferentes variâncias, e para simular da distribuição *t* é interessante conhecer sua representação estocástica. Seja $\mathbf{Z}^* \sim N_d(\mathbf{0}, \boldsymbol{\Sigma})$ o vetor aleatório com distribuição Normal d -dimensional com médias zero e matriz de covariâncias $\boldsymbol{\Sigma}$. O vetor aleatório \mathbf{r} , $\mathbf{r} \sim t_d(\nu, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, possui a representação

$$\mathbf{r} \stackrel{d}{=} \boldsymbol{\mu} + \sqrt{W}\mathbf{Z}^*, \quad (2.14)$$

onde W é uma v.a. independente de \mathbf{Z}^* tal que $\frac{\nu}{W} \sim \chi_\nu^2$, isto é, W segue uma distribuição Gama Inversa, $W \sim \text{Ig}(\nu/2, \nu/2)$ (veja Fang, K., Kotz, S. e Ng, W. (1990)).

Alternativamente (2.14) pode ser escrita como

$$\mathbf{r} \stackrel{d}{=} \boldsymbol{\mu} + \sqrt{W}\mathbf{A}\mathbf{Z},$$

onde $\mathbf{Z} \sim N_d(\mathbf{0}, \mathbf{I}_d)$, e \mathbf{A} é uma matriz $d \times d$ de constantes. Condiçãoando na v.a. W temos que $\mathbf{r}|W = w \sim N_d(\boldsymbol{\mu}, w\boldsymbol{\Sigma})$, onde $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}'$. Fica claro que neste caso a distribuição de \mathbf{r} é construída a partir de distribuições Normais multivariadas com o mesmo vetor de médias e matrizes de covariâncias que diferem apenas quanto a um fator multiplicativo w . Esse conjunto de valores w são geralmente entendidos como “pesos” fornecidos pela distribuição de W . A v.a. W em finanças pode representar os choques da economia que produzem mudanças na volatilidade.

Pertencendo à classe de distribuições elípticas, os pontos para os quais a densidade (2.13) possui o mesmo valor formam um elipsoide dado pela equação $(\mathbf{r} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{r} - \boldsymbol{\mu}) = c$, para algum $c > 0$. Sugerimos ao leitor fazer, através de uma figura, uma comparação no caso bivariado entre a distribuição Normal bivariada com $\rho = 0.6$ e a distribuição *t*-bivariada com $\nu = 4$ e $\rho = 0.6$. Ambas distribuições possuem $\boldsymbol{\mu} = \mathbf{0}$ e a mesma matriz de covariâncias. Poderia grafar as linhas de contorno *possuindo o mesmo valor de densidade* nos dois casos. Observará que os contornos de igual densidade decaem mais lentamente nas “caudas” da distribuição *t* aumentando as chances de ocorrerem eventos extremos simultâneos. No plano univariado observamos caudas mais pesadas para distribuição *t* quando comparadas à Normal.

A Tabela 2.3 fornece as estimativas de máxima verossimilhança da locação, do desvio padrão, e dos graus de liberdade da *t* para os quatro índices. A tabela também mostra os estimadores clássicos, os

EMV sob normalidade. É muito intrigante ver como a média amostral fornece estimativa tão diferente da t -simétrica para a locação dos retornos. O teste GOF de Kolmogorov aceitou a hipótese nula de boa qualidade do ajuste t .

Tabela 2.3: *Estimativas de máxima verossimilhança dos parâmetros μ , σ , λ , e ν , sob os modelos Normal, t -simétrico e t -assimétrico.*

Estimadores	Brasil	Mexico	Indonésia	China
Média Amostral	0.0417	0.0214	0.0770	0.0431
D.P. Amostral	2.7954	2.1405	2.0517	2.2539
GOF-Normal(D_T)	Rj. H_0	Rj. H_0	Rj. H_0	Rj. H_0
t -Simétrica $\hat{\mu}$	0.1753	0.1004	0.1467	0.1174
t -Simétrica $\hat{\nu}$	2.9433	2.8601	2.7877	2.9193
t -Simétrica $\widehat{D.P.}$	3.0377	2.3712	2.3085	2.5129
GOF- $t(D_T)$	Ac. H_0	Ac. H_0	Ac. H_0	Ac. H_0
t -Assimétrica $\hat{\mu}$	0.0398	-0.0088	0.0737	0.0336
t -Assimétrica $\hat{\lambda}$	-0.1035	-0.1032	-0.0709	-0.0724
t -Assimétrica $\hat{\nu}$	2.9913	2.8484	2.8218	2.9573
t -Assimétrica $\widehat{D.P.}$	3.0183	2.3899	2.2900	2.4956
GOF- t -Assimétrica(D_T)	Ac. H_0	Ac. H_0	Ac. H_0	Ac. H_0

D.P. significa desvio padrão, nos três casos corresponde ao parâmetro σ

2.8 Distribuições Elípticas Assimétricas

Uma grande classe de misturas de distribuições Normais multivariadas, obtidas a partir de misturas com diferentes médias e variâncias, pode ser obtida generalizando-se (2.14). Temos então

$$\mathbf{r} = \boldsymbol{\mu} + \gamma g(W) + \sqrt{W} \mathbf{Z}^* \quad (2.15)$$

para alguma função $g : [0, \infty) \rightarrow [0, \infty)$ e um d -dimensional vetor de parâmetros γ . Quando $\gamma \neq \mathbf{0}$ (2.15) representa uma família assimétrica sem as propriedades de simetria de uma família elíptica. Para outras definições de distribuição elíptica assimétrica multivariada veja Genton (2004).

Quando $g(W) = W$ e $W \sim Ig(\nu/2, \nu/2)$ a mistura resultante segue uma distribuição t multivariada assimétrica, embora este nome também seja usado em outras definições. Neste caso a densidade é dada por

$$f(\mathbf{r}) = c \frac{\Gamma(\frac{\nu+d}{2}) (\sqrt{(\nu + \mathbf{r} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{r} - \boldsymbol{\mu}) \gamma' \boldsymbol{\Sigma}^{-1} \gamma}) \exp(\mathbf{r} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} \gamma)}{(\sqrt{(\nu + \mathbf{r} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{r} - \boldsymbol{\mu}) \gamma' \boldsymbol{\Sigma}^{-1} \gamma})^{-\frac{\nu+d}{2}}} \left(1 + \frac{(\mathbf{r} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{r} - \boldsymbol{\mu})}{\nu} \right)^{-\frac{\nu+d}{2}} \quad (2.16)$$

onde K_λ denota a função de Bessel de terceiro tipo, veja Abramowitz, M. e Stegun, I. (1965), capítulos 9 e 10) e onde a constante normalizadora é

$$c = \frac{2^{\frac{2-(\nu+d)}{2}}}{\Gamma(\frac{\nu}{2})(\pi\nu)^{d/2}|\Sigma|^{1/2}}.$$

Usaremos a notação $\mathbf{r} \sim t_d(\nu, \boldsymbol{\mu}, \Sigma, \lambda)$ para representar esta distribuição. A partir das propriedades de K_λ pode-se mostrar que a densidade da t -assimétrica converge para a densidade da distribuição t simétrica usual quando $\lambda \rightarrow 0$. Os momentos desta distribuição podem ser facilmente calculados devido a sua estrutura de mistura de Normais. São dados por

$$E[\mathbf{r}] = E[E(\mathbf{r} \mid W)] = \boldsymbol{\mu} + E[W]\gamma = \boldsymbol{\mu} + \frac{\nu}{\nu-2}\gamma$$

$$\text{cov}(\mathbf{r}) = E[\text{var}(\mathbf{r} \mid W)] + \text{var}(E[\mathbf{r} \mid W]) = \frac{\nu}{\nu-2}\Sigma + \frac{2\nu^2}{(\nu-2)^2(\nu-4)}\gamma\gamma'.$$

As covariâncias desta mistura de médias e variâncias são finitas para $\nu > 4$ (para a t -simétrica apenas é preciso $\nu > 2$). Isto faz com que as distribuições marginais possuam caudas ainda mais pesadas do que o caso onde $\lambda = 0$.

2.9 Distribuição t Assimétrica

Simetria elíptica implica que todas as distribuições marginais univariadas sejam simétricas, fato que *não é* na maioria das vezes empiricamente verificado para retornos financeiros. Por isto, nesta seção daremos uma atenção especial à distribuição t assimétrica que é capaz de lidar com duas características usualmente observadas em retornos, assimetria e caudas pesadas. Assim, na modelagem de retornos financeiros, a distribuição t assimétrica tem papel muito importante, sendo capaz de eficientemente representar os dados, veja por exemplo Fantazzini, D. (2007) e Mendes, B. V. M., Semeraro, M. M. e Leal, R. P. C. (2010).

Ambas as referências acima utilizam uma versão da t assimétrica proposta por Hansen, B. (1994), uma família flexível que permite calibrar a locação (μ), a escala (σ), a cauda pesada (ν), e a assimetria (λ), mantendo ainda a simplicidade da sua versão simétrica. Quando $\lambda = 0$ temos o caso simétrico.

A densidade da t -assimétrica tem forma fechada e a implementação do método da máxima verossimilhança é viável pois são apenas 4 parâmetros a estimar. A moda da distribuição é menor (maior) que μ se a assimetria é para a direita (esquerda). A Figura 2.3 mostra a densidade da t assimétrica de Hansen para $\nu = 4$ e $\lambda = -0.6, 0.0, 0.6$.

A f.d.a. da t assimétrica com $\mu = 0$ e $\sigma = 1$ (veja Fantazzini, D. (2007)) é dada por

$$G(y; \nu, \lambda) \begin{cases} (1 - \lambda)G_T\left(\sqrt{\frac{\nu}{\nu-2}}\left(\frac{by+a}{1-\lambda}\right); \nu\right), & \text{para } y < -\frac{a}{b} \\ (1 - \lambda)/2 & \text{para } y = -\frac{a}{b} \\ (1 + \lambda)G_T\left(\sqrt{\frac{\nu}{\nu-2}}\left(\frac{by+a}{1+\lambda}\right); \nu\right) - \lambda, & \text{para } y > -\frac{a}{b} \end{cases} \quad (2.17)$$

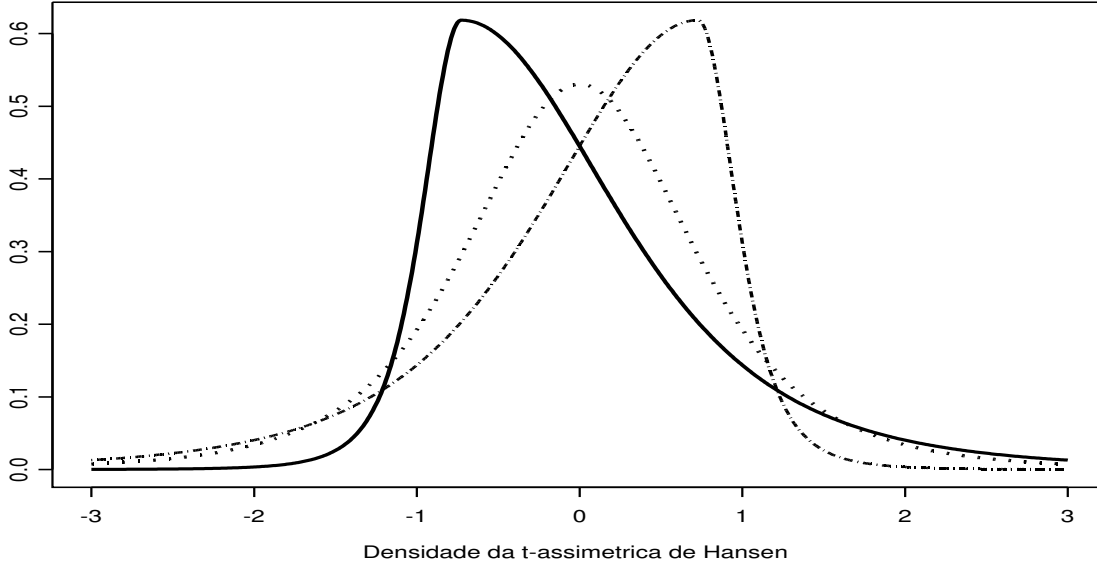


Figura 2.3: Densidade da t assimétrica com $\mu = 0$, $\sigma = 1$, $\nu = 4$, e $\lambda = -0.6, 0.0, 0.6$, respectivamente representadas pelas linhas sólida, pontilhada, e tracejada.

onde $G_T(t; \nu)$ representa a f.d.a. da distribuição t -Student (simétrica) com ν graus de liberdade, onde

$$c = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\nu/2)\sqrt{\pi(\nu-2)}}$$

$$b = \sqrt{1 + 3\lambda^2 - a^2}$$

$$a = 4\lambda c \left(\frac{\nu-2}{\nu-1} \right).$$

As estimativas de máxima verossimilhança sob o modelo t assimétrico de Hansen são também dadas na Tabela 2.3. É novamente intrigante ver como a média amostral e a t assimétrica fornecem estimativas parecidas e tão diferentes da t -simétrica para a locação dos retornos. Entretanto, o gráfico das duas densidades t estimadas sobre o histograma para o Brasil, por exemplo, são visualmente coincidentes. O teste GOF de Kolmogorov aceitou a hipótese nula de boa qualidade do ajuste t assimétrico para os 4 índices.

Outras duas propostas para a t assimétrica que não apresentam dificuldade de implementação são Jones, M.C. e Faddy, M.J. (2003) e Dongming Zhua e John W. Galbraith (2010).

2.10 Valores Extremos

A teoria dos valores extremos (TVE) tem estado presente na literatura por um tempo já relativamente longo. A preocupação com a modelagem de eventos extremos data do século XVII em estudos em astronomia (veja em Gumbel, E. J. (1958)). Alguns autores colocam o artigo de Bortkiewicz, L. von. (1922) como um marco no desenvolvimento formal da TVE. Neste artigo, Bortkiewicz trata da distribuição do tamanho do intervalo entre o máximo e o mínimo numa amostra da distribuição normal. Resultados importantes no caso univariado foram obtidos em 1928 por Fisher e Tippett. Avanços no caso bivariado foram obtidos ao final da década de 50 e no início da de 60 (Gumbel, E. J. (1954), Gumbel, E. J. (1958), Oliveira, T. (1958), Oliveira, T. (1962,1963), Sibuya, M. (1960)). Durante as últimas duas décadas vimos várias técnicas estatísticas, desenvolvidas sob a denominação geral de teoria dos valores extremos, experimentarem uma aceitação prática crescente, especialmente nas áreas de finanças e atuária. Esta popularidade é muito devida à sua habilidade de melhor quantificar as probabilidades de ocorrência de eventos raros, ou mesmo nunca antes observados, de um processo.

A TVE é um ramo da probabilidade que estuda o comportamento estocástico de extremos associados a um conjunto de variáveis aleatórias (ou vetores aleatórios) com distribuição comum F . Dentro da denominação geral de extremos incluímos o máximo e o mínimo, estatísticas de ordem extremas e excessos além de um limiar. As características e propriedades das distribuições desses extremos aleatórios são determinadas pelas caudas (inferior e superior, no caso univariado) da distribuição subjacente F . Restringir a atenção às caudas de uma distribuição apresenta a vantagem de termos a nossa disposição diversos modelos estatísticos adequados para as mesmas. Esses modelos nos permitirão fazer inferências mais precisas sobre as caudas da distribuição de interesse F .

Fréchet, M. (1927) introduziu a distribuição assintótica do máximo de uma amostra. Fisher, R. A. e Tippett, L. H. C. (1928) estudando independentemente o mesmo problema, chegaram ao resultado que estabelece três tipos possíveis para a distribuição do máximo. Gnedenko, B. V. (1943) introduziu a distribuição de valores extremos generalizada e forneceu as condições necessárias e suficientes para a convergência em distribuição do máximo e do mínimo.

Vários artigos contendo aplicações dos resultados acima se seguiram nas décadas de 30 e 40. Estas aplicações envolviam, na maioria das vezes, análises de inundações, níveis de chuva, resistência de materiais etc. Na década de 50, um acontecimento marcante², relatado em Embrechts, P., Klüppelberg, C. e Mikosch, T. (1997), fez com que o governo alemão estabelecesse o comitê Delta encarregado de modelar estatisticamente eventos extremos de forma a prevenir contra inundações catastróficas futuras. Hoje existem numerosos estudos e aplicações da TVE espalhados em artigos e livros. Impossível listá-los todos aqui. Mencionamos apenas as contribuições de Leadbetter, M., Lindgren, G. e Rootzén, H. (1983), Embrechts, P., Klüppelberg, C. e Mikosch, T. (1997), Coles, S. (2001), e Reiss, R-D. e Thomas, M. (1997). Em finanças destacamos os livros Mendes, B. V. M. (2004) e McNeil, A. J., Frey, R. e Embrechts, P. (2005), e vários artigos como Smith, R. L. (1987), Smith, R.L. (1999), Longin, F. (1996), Danielsson, J. e De Vries, C.G. (1997), McNeil, A.J. (1998), Mendes, B. V. M. (2000b), McNeil, A. J. e Frey, R. (2000), entre outros .

²Em 1º de fevereiro de 1953, diques em várias localidades da região da Holanda foram ultrapassados pelas águas como consequência de uma grande tempestade, ocasionando a morte de 1.800 pessoas. Este foi um evento extremo para o qual o sistema de proteção não estava preparado.

Modelagem para os Máximos

Conforme comentamos no Capítulo 1, grandes quedas em mercados financeiros podem não ser respostas imediatas a acontecimentos ou notícias de grande impacto. Isto é, a ocorrência ocasional de movimentos extremos de um processo pode ser prevista usando a informação fornecida pelo processo gerador dos retornos. Assim, utilizando o conjunto de técnicas bem fundamentadas da TVE, o gerente de riscos financeiros pode usar da melhor forma possível a pouca informação que tem sobre movimentos extremos de mercado. Por exemplo, considere o comportamento do índice do Brasil na Figura 2.4. Nesta figura vemos os movimentos dos preços e dos retornos diários deste índice de 4 de janeiro de 2006 a 4 de julho de 2008. Aparentemente não há nada que nos faça suspeitar que alguma grande queda está por vir. Imaginemos agora que nesta sexta-feira, dia 4 de julho de 2008, lhe fosse pedido para prever um cenário adverso para os próximos meses. Usando a modelagem para os mínimos da TVE, você decide ser bem cauteloso e calcula o evento de 48 meses, um valor de retorno que se espera ser excedido uma vez a cada 4 anos. A metodologia lhe fornece uma estimativa pontual de $-10,66\%$ e um intervalo de 95% de confiança igual a $[-19,27\%, -8,31\%]$. Vale a pena lembrar que nos três meses seguintes este índice apresentou as quedas catastróficas de $-18,32\%$ (15/Out), $-17,07\%$ (22/Out), $-14,39\%$ (3/Out), e $-13,60\%$ (29/Set) entre outras. Extremos de magnitude não observada na série histórica, todos dentro do intervalo de confiança.

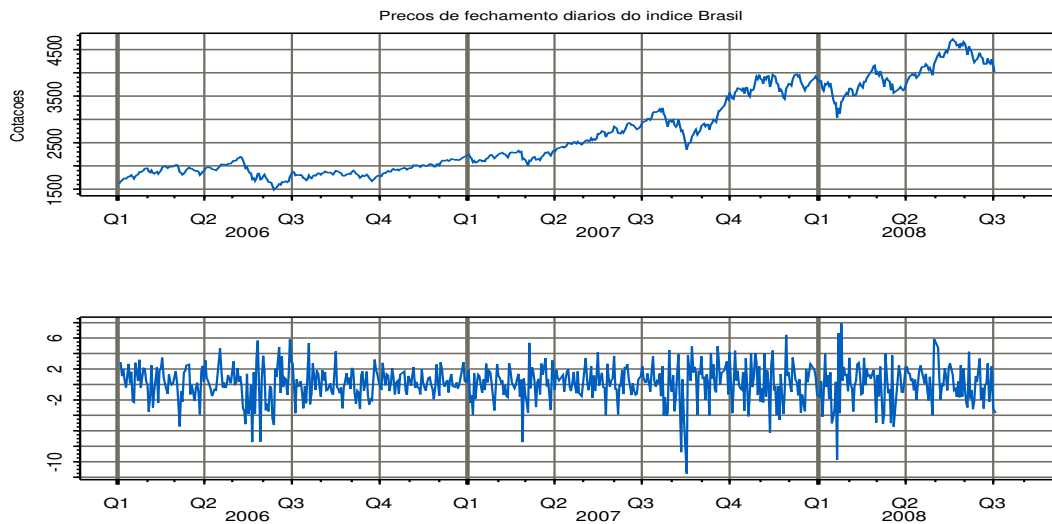


Figura 2.4: Índice Brasil de 4 de janeiro de 2006 a 4 de julho de 2008.

A boa qualidade das estimativas deste exemplo deve-se ao fato que focamos apenas nas caudas de F . Sendo os eventos extremos os eventos atípicos, diferentes da grande maioria das outras observações,

e ocorrendo com probabilidade baixa, a estimação das probabilidades associadas aos mesmos torna-se mais difícil devido à escassez de dados. Um dos principais resultados da TVE, o Teorema de Fisher-Tippett (Fisher, R. A. e Tippett, L. H. C., 1928), fornece o resultado de convergência fraca para o máximo centrado e normalizado e vem resolver este problema.

Seja $\{r_t\}$, $t = 1, 2, \dots, n$, uma sequência de v.a.'s i.i.d. com distribuição F , e seja M_n o máximo dos valores r_t no bloco de tamanho n , isto é, $M_n = \max_{1 \leq t \leq n} \{r_t\}$. Temos que

$$Pr\{M_n \leq x\} = (F(x))^n.$$

Seja x^+ o limite à direita do suporte de F . Sendo $x^+ < +\infty$, quando $n \rightarrow \infty$, $(F(x))^n$ converge para zero ou é igual a um, dependendo se $x < x^+$ ou $x \geq x^+$, resultado este que não é útil para modelagens.

Por outro lado, para conhecer-se a função de distribuição exata do máximo M_n para n pequeno ou moderado, precisaríamos especificar a distribuição subjacente F , desconhecida e objeto da pesquisa. Para n moderado aproximamos esta distribuição pelo resultado assintótico. O Teorema de Fisher-Tippett estabelece que se existirem sequências de constantes reais $\{a_n\}$ e $\{b_n\}$ com $a_n > 0$ tais que

$$P\left\{\frac{M_n - b_n}{a_n} \leq z\right\} \rightarrow G(z) \text{ quando } n \rightarrow \infty,$$

onde G é uma função de distribuição não-degenerada, então G é dada por

$$G(z) = \exp\left(-\left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right), \quad (2.18)$$

definida no conjunto $\{z : 1 + \xi(z - \mu)/\sigma > 0\}$, e onde os parâmetros satisfazem $-\infty < \mu < \infty, \sigma > 0$ e $-\infty < \xi < \infty$.

Essa é a distribuição de Valores Extremos Generalizada (*Generalized Extreme Value* (GEV)), introduzida por von Mises (1964) e Jenkinson, A. F. (1955), e caracterizada pelos parâmetros de localização (μ), escala (σ) e forma (ξ). Essa distribuição tem como casos particulares as distribuições Gumbel (tipo I, $\xi = 0$), Weibull (tipo III, $\xi < 0$, $x^+ < \infty$) e Fréchet (tipo II, $\xi > 0$, $x^+ = \infty$). O parâmetro ξ ou seu recíproco $\frac{1}{\xi}$ é chamado de índice de cauda.

Veja na Figura 2.5 as densidades dessas três distribuições nas suas formas padrões. O caso $\xi = 0$ é interpretado como o limite de (2.18) quando $\xi \rightarrow 0$, conduzindo à função de distribuição

$$G(z) = \exp\left(-\exp\left[-\left(\frac{z - \mu}{\sigma}\right)\right]\right), -\infty < z < \infty.$$

Se o máximo (padronizado) de uma sequência de tamanho n com distribuição subjacente F segue uma distribuição GEV, então dizemos que F pertence ao domínio de atração da GEV, para um de seus três tipos. Isto é, se

$$|F^n(x) - G(\frac{x - b_n}{a_n})| \rightarrow 0, \quad n \rightarrow \infty \quad (2.19)$$

se verifica para algum dos tipos da G , dizemos que F pertence ao domínio de atração do máximo da distribuição de valores extremos G , e denotamos por $F \in MDA(G)$. Por exemplo, a distribuição t -Student pertence ao domínio de atração da tipo II Fréchet, $MDA(G_{II})$.

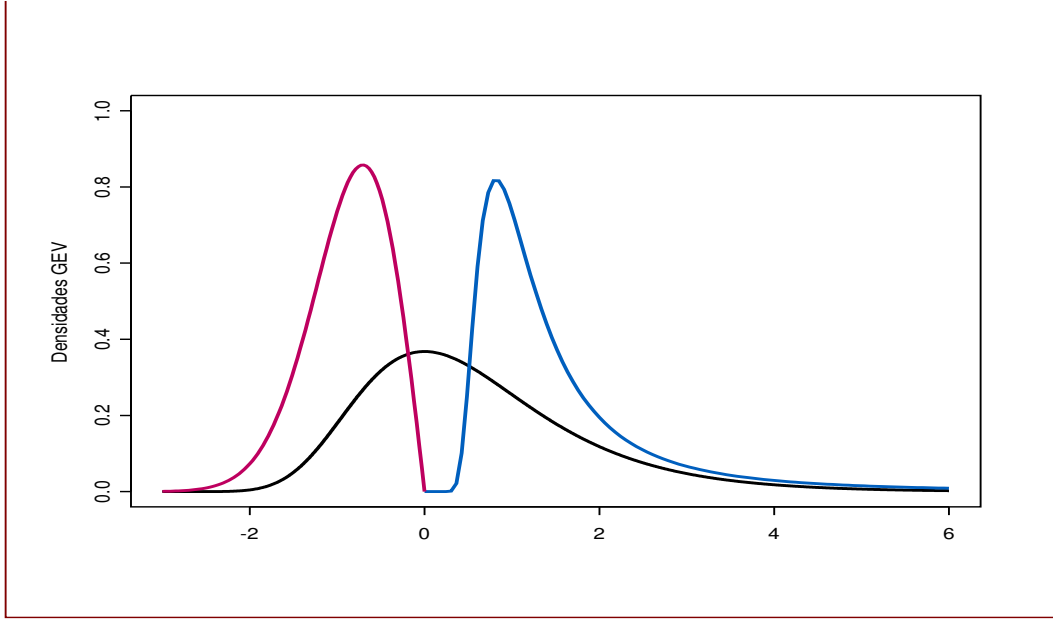


Figura 2.5: Densidades dos três tipos de distribuições de valores extremos nas suas formas padrões. Gumbel em preto, Weibull em rosa e Fréchet em azul.

Um resultado bastante conhecido devido a Gnedenko é que $F \in MDA(G_{II})$ se e somente se a cauda $\bar{F}(r) = 1 - F(r)$ varia regularmente no infinito com parâmetro $-1/\xi < 0$, isto é,

$$\bar{F}(r) = L(r)r^{-1/\xi}, \quad r > 0,$$

onde a função $L(r)$ possui variação lenta no infinito, isto é, $\lim_{r \rightarrow \infty} \frac{L(sr)}{L(r)} = 1$, $s > 0$ (mais detalhes no Apêndice [AP2.4]).

Para cada $F \in MDA(G)$, a qualidade e velocidade da aproximação da distribuição do máximo padronizado para G depende de F . As provas de convergência se baseiam em convergência de momentos, convergência de densidades, convergência da f.d.a. etc. Por exemplo, o máximo de uma exponencial converge mais rápido para a Gumbel do que o de uma Normal. A Figura 2.6 ilustra esse fato. Esta figura compara as distribuições Gumbel e a distribuição exata do máximo de uma exponencial unitária, para $n = 30$. Observa-se uma excelente aproximação. Para melhor apreciar esse resultado o leitor poderia compará-lo com outro obtido substituindo-se a exponencial unitária pela normal-padrão (veja exercício no final deste capítulo).

É possível caracterizar o domínio de atração do máximo (a coleção de F 's) para cada tipo de distribuição de valores extremos. Por exemplo, von Mises fornece condições suficientes para F pertencer ao $MDA(G)$ para alguma G , no caso de F ter densidade. É o conceito de equivalência de cauda, o qual define uma relação de equivalência para as F 's no mesmo domínio de atração do máximo.

Duas distribuições F e H são ditas equivalentes de cauda se elas possuem o mesmo limite superior x^+ , e $\lim_{x \uparrow x^+} \bar{F}(x)/\bar{H}(x) = c$, para alguma constante $0 < c < \infty$. É possível mostrar que se duas

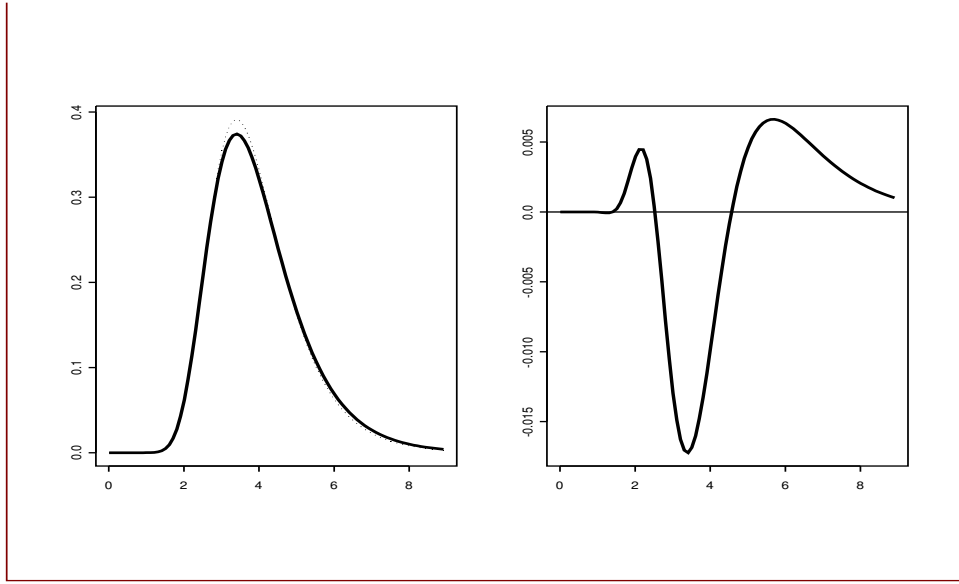


Figura 2.6: Esquerda: Densidade f da f.d.a $(1 - e^{-x})^{30}$ (linha contínua) e densidade Gumbel g (linha pontilhada) para $\mu = 3.4$ e $\sigma = 0.94$. Direita: Diferença $f - g$.

distribuições F e H são equivalentes de cauda então pertencem ao mesmo domínio de atração (veja Embrechts, P., Klüppelberg, C. e Mikosch, T. (1997)). Além disto, elas admitem as mesmas constantes normalizadoras para o máximo.

de Haan, L. (1976) dá duas condições necessárias para $F \in MDA(G)$ para alguma G . Colocando de uma forma simples, essas condições estabelecem que: (1) Se $F \in MDA(G_I)$ e $x^+ = \infty$, então a cauda de F decai exponencialmente e todos os momentos de F existem. (2) Se $F \in MDA(G_{II})$ então $x^+ = \infty$, F tem caudas pesadas que decaem por uma potência, e nem todos os momentos são finitos (depende do valor de ξ).

Jansen, D.W. e De Vries, C.G. (1991) dão uma condição suficiente para $F \in MDA(G_{II})$. Esta condição é conhecida como variação regular no infinito e é relacionada com um resultado de Feller, W. (1971), cap. VIII. Esta condição estabelece que para $F \in MDA(G_{II})$ é suficiente que $x^+ = \infty$ e que para todo $x > 0$ e algum $\xi > 0$, tenha-se

$$\lim_{x \rightarrow \infty} \frac{1 - F(xt)}{1 - F(x)} = t^{-\alpha} \quad t > 0.$$

Em outras palavras, o resultado limite estabelece que para algumas F com $x^+ = \infty$ e tais que $\bar{F}(x) = x^{-\alpha}L(x)$, onde $L(x)$ é uma função de variação lenta, o máximo M_n coletado em uma amostra i.i.d. de tamanho n satisfaz: $c_n^{-1} M_n \xrightarrow{d} G_{II}$. A seguir daremos alguns exemplos que ajudam a compreender e ligar as várias possibilidades para a distribuição F com os três domínios de atração do máximo.

1) A Gumbel como distribuição-limite do máximo: Para a distribuição Gumbel padrão, a constante normalizadora do máximo a_n é função de b_n e $b_n = F^{\leftarrow}(1 - 1/n)$, onde F^{\leftarrow} representa a inversa gene-

realizada (veja definição no Apêndice [AP2.6] deste capítulo). Para F com $x^+ \leq \infty$ e tais que $\bar{F}(x) = c(x)\exp\{-\int_{x_0}^x \frac{g(t)}{a(t)}dt\}$, $x_0 < x < x^+$, onde, quando $x \uparrow x^+$, $c(x) \rightarrow c > 0$, $g(x) \rightarrow 1$, $a'(x) \rightarrow 0$, temos o resultado limite. Como exemplos de $F \in MDA(G_I)$ temos: (i) Normal com $a_n = (2\ln(n))^{-1/2}$ e $b_n = \sqrt{2\ln(n)} - \frac{\ln(4\pi) + \ln(\ln(n))}{2(2\ln(n))^{1/2}}$; (ii) exponencial; (iii) gama; misturas discretas de Normais. O domínio de atração da Gumbel não é facilmente caracterizável. As distribuições F neste domínio possuem cauda direita que decai mais rápido do que qualquer potência. Isto inclui distribuições com caudas leves como a Normal e com caudas moderadamente pesadas como a Lognormal.

2) A Fréchet como distribuição-limite de máximo: Para a distribuição Fréchet padrão, as constantes normalizadoras do máximo são $a_n = F^{\leftarrow}(1 - 1/n)$ e $b_n = 0$. Exemplos de F satisfazendo a condição dada acima são: (i) Cauchy padrão com $a_n = n/\pi$; (ii) t-student(ν); (iii) Loggama (veja exemplo 3.3.11 de Embrechts, P., Klüppelberg, C. e Mikosch, T. (1997)); (iv) processos ARCH (prova em de Haan, L., Resnick S. I., Rootzén, H. e de Vries, C. G. (1989); ver definição de processo ARCH no Capítulo 4). No caso da t-student(ν), o número de graus de liberdade ν é igual ao índice de cauda α . Tem-se também que soma de um número finito de t-student(ν) pertence ao domínio de atração da Fréchet, com o mesmo α . Isto explica uma certa estabilidade da estimativa de $\alpha = 1/\xi$ quando se varia o tamanho n do bloco. A distribuição Fréchet é a que mais interessa em finanças, já que esta é a distribuição-limite de máximos coletados de distribuições com caudas pesadas e com $x^+ = \infty$. De fato, as distribuições F no domínio de atração de G_{II} são tais que $E(X^+)^{\delta} = \infty$ para $\delta > \alpha$, onde X^+ representa a parte positiva de X . Em geral retornos financeiros possuem caudas pesadas e são, a princípio, ilimitados.

3) A Weibull como distribuição-limite do máximo: Para a distribuição Weibull padrão, as constantes normalizadoras do máximo são $a_n = x^+ - F^{\leftarrow}(1 - 1/n)$ e $b_n = 0$. O resultado limite estabelece que $a_n^{-1}(M_n - x^+) \xrightarrow{d} G_{III}$ e depende de $x^+ < \infty$ e de que $\bar{F}(x^+ - 1/x) = x^{-\alpha}L(x)$. Exemplos de F no $MDA(G_{III})$ são: (i) Uniforme(0,1), com $a_n = \text{constante} \cdot n^{-1/\alpha}$, $b_n = x^+$, (ii) Beta.

Para maiores detalhes e provas formais dos resultados relacionando as distribuições F , suas constantes normalizadoras, e os domínios de atração do máximo, veja a seção 3.3 de Embrechts, P., Klüppelberg, C. e Mikosch, T. (1997).

Em essência, existem similaridades entre o teorema de Fisher e Tippet e o teorema central do limite. O teorema central do limite estabelece que dentro das f.d. não degeneradas apenas as distribuições estáveis podem ser distribuições limite. O teorema de Fisher e Tippet estabelece que dentro das f.d. não degeneradas apenas as distribuições max-estáveis podem ser distribuições limite. Outras relações probabilísticas interessantes entre F e G podem ser vistas em Embrechts, P., Klüppelberg, C. e Mikosch, T. (1997).

Retornos financeiros em geral apresentam conglomerados (*clusters*) de valores extremos, violando a suposição de independência. No caso da sequência não ser i.i.d., uma extensão do teorema de Fisher e Tippet para séries estacionárias estabelece que

$$Pr\{M_n \leq x\} = (F(x))^{\theta n}, \quad (2.20)$$

onde θ é o índice extremal, $0 \leq \theta \leq 1$, sendo igual a um no caso particular de v.a.'s independentes. Em outras palavras, temos que o máximo de n/θ v.a.'s *dependentes* tem aproximadamente a mesma

distribuição ($Pr\{M_{n/\theta} \leq x\}$) do máximo de n v.a.'s i.i.d. Para maiores detalhes veja Leadbetter, M., Lindgren, G. e Rootzén, H. (1983).

Na prática, dada uma série de tamanho T de retornos $\{r_1, \dots, r_T\}$, para a estimação do modelo GEV iremos sempre considerar a cauda direita (os máximos). Quando o interesse for na cauda esquerda multiplica-se a série por (-1) . Fixamos n e coletamos as m observações de máximos (m_1, m_2, \dots, m_m) , os máximos de cada um dos blocos de tamanho n , e supomos que esta seja uma amostra i.i.d. dos máximos (ou dos valores absolutos de mínimos). De fato, n deve ser escolhido de tal forma que essa suposição seja satisfeita.

As estimativas dos parâmetros ξ , μ e σ da distribuição $G_{\xi, \mu, \sigma}$ podem ser obtidas por vários procedimentos estatísticos, incluindo o Método dos Momentos (Reiss, R. D. (1997)), Método da Regressão (Reiss, R. D. (1997)), Método da Máxima Verossimilhança (Embrechts, P., Klüppelberg, C. e Mikosch, T., 1997), e o Método dos L-momentos (Hosking, J. R. M., 1990). Veremos aqui os estimadores de máxima verossimilhança (EMV) e os de L-momentos, por suas boas propriedades e fácil implementação.

Os EMV de ξ , μ e σ são obtidos numericamente pela maximização da função de verossimilhança, isto é, são os valores em $\mathbb{R} \times \mathbb{R} \times \mathbb{R}^+$ que maximizam

$$\mathcal{L}(\xi, \mu, \sigma; (m_1, \dots, m_m)) = \log \prod_{i=1}^m g_{\xi, \mu, \sigma}(m_i) I_{\{1 + \frac{\xi}{\sigma}(m_i - \mu) > 0\}} , \quad (2.21)$$

onde $g_{\xi, \mu, \sigma}(\cdot)$ representa a densidade de $G_{\xi, \mu, \sigma}$.

Os EMV podem ser obtidos resolvendo-se através de um processo iterativo o sistema de equações não lineares obtido a partir da diferenciação de (2.21) em relação à ξ , μ , e σ . Este procedimento, contudo, pode tornar-se algumas vezes numericamente instável. Nos casos regulares, os estimadores de máxima verossimilhança são consistentes, eficientes (sob o verdadeiro modelo), e assintoticamente Normais. Os casos não regulares ocorrem quando o suporte da distribuição depende dos parâmetros desconhecidos, como pode acontecer no caso da GEV. Smith, R. L. (1985) fornece detalhes teóricos sobre o comportamento assintótico dos EMV dos parâmetros da GEV e mostra que os mesmos mantêm suas boas propriedades sempre que $\xi > -\frac{1}{2}$. Este fato é importante para as aplicações que fazemos aqui, pois as distribuições F com caudas pesadas geralmente pertencem ao $MDA(G_{II})$.

O método dos L-momentos é similar ao método do Momentos (veja Apêndice do Capítulo 2 [AP2.6]). Os L-momentos são modificações dos “probability weighted moments” (PWM), de Greenwood, J. A., Landwehr, J. M., MAtalas, N. C. e Wallis, J. R. (1979). Os PWM de uma v.a. Y são definidos como

$$\eta_{p,q,s} = E[Y^p (F(Y))^q (1 - F(Y))^s] .$$

Especialmente úteis são os obtidos no caso particular:

$$\eta_{1,q,0} = \beta_q = E[Y (F(Y))^q] .$$

Os PWM têm sido usados como base para métodos de estimação de parâmetros de uma f.d.a., mas são difíceis de serem interpretados como ou de serem relacionados diretamente com os parâmetros de escala e forma da distribuição. Informações sobre estas e outras características de uma distribuição

são melhor obtidas usando-se combinações lineares dos PWM. Por exemplo, $2\beta_1 - \beta_0$ e $6\beta_2 - 6\beta_1 + \beta_0$ representam, respectivamente, a escala e a assimetria de uma distribuição.

Os L-momentos λ_{q+1} , $q = 0, 1, 2, \dots$ são definidos (ver Hosking, J. R. M. (1990)) para v.a.s com esperança finita, como o valor esperado de combinações lineares dos dados e podem ser escritos como

$$\lambda_{q+1} = \sum_{j=0}^q p_{q,j}^* \beta_j, \quad q = 0, 1, 2, \dots \quad (2.22)$$

$$p_{q,j}^* = \frac{(-1)^{q-j} (q+j)!}{(j!)^2 (q-j)!} \quad (2.23)$$

O L-momento relacionado com a locação da distribuição é λ_1 . O L-momento relacionado com a escala é λ_2 ($\lambda_2 = 2\beta_1 - \beta_0$) e deve ser comparado com o desvio-padrão σ . Tem-se que $\sigma \geq \sqrt{3}\lambda_2$. O L-momento representando a assimetria é $\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0$.

Para uma amostra de tamanho m , os PWM amostrais b_q de ordem q , $q = 0, 1, \dots, m-1$, são estimadores não viciados dos PWM populacionais β_q , e podem ser calculados usando

$$b_q = \frac{1}{m} \sum_{j=q+1}^m \frac{(j-1)(j-2)\dots(j-q)}{(m-1)(m-2)\dots(m-q)} m_{(j)},$$

com $b_0 = \frac{1}{m} \sum_{j=1}^m m_{(j)}$.

Similarmente a (2.22) os L-momentos amostrais são definidos como

$$l_{q+1} = \sum_{j=0}^q p_{q,j}^* b_j, \quad q = 0, 1, \dots, m-1, \quad (2.24)$$

onde $p_{q,j}^*$ são definidos em (2.23). Os L-momentos amostrais l_q são estimadores não viciados de λ_q .

O método dos L-momentos consiste em equacionar um número fixo de L-momentos amostrais às quantidades populacionais correspondentes. Assim, expressamos os parâmetros de uma distribuição em função dos L-momentos. Para a GEV essas equações são

$$\lambda_1 = \mu - \frac{\sigma}{\xi} (1 - \Gamma(1 - \xi))$$

$$\lambda_2 = -\frac{\sigma}{\xi} (1 - 2^\xi) \Gamma(1 - \xi)$$

$$\frac{\lambda_3}{\lambda_2} = 2 \frac{(1-3^\xi)}{(1-2^\xi)} - 3$$

onde $\Gamma(\cdot)$ denota a função Gama.

Para estimar os parâmetros μ , σ e ξ substituímos nas equações acima os L-momentos populacionais pelos seus correspondentes amostrais l_1 , l_2 e l_3 . Para estimar ξ resolvemos a terceira equação. Não existe solução explícita, mas em Hosking, J. R. M. e Wallis, J. R. (1997) temos a seguinte aproximação: $\hat{\xi} = -7.8590c - 2.9554c^2$, onde $c = \frac{2}{3+l_3/l_2} - \frac{\log 2}{\log 3}$. Em seguida os outros dois parâmetros podem ser obtidos explicitamente a partir das expressões:

$$\hat{\sigma} = \frac{-l_2 \hat{\xi}}{(1 - 2^{\hat{\xi}}) \Gamma(1 - \hat{\xi})},$$

$$\hat{\mu} = l_1 + \frac{\hat{\sigma}}{\hat{\xi}}(1 - \Gamma(1 - \hat{\xi})) .$$

De acordo com Hosking, J. R. M. e Wallis, J. R. (1997), os estimadores de L-momentos possuem vício e variância pequenos e a propriedade de unicidade, além de serem mais robustos em relação à outliers do que os EMV. Para amostras de tamanho pequeno ou moderado os L-momentos são geralmente mais eficientes do que os EMV. Sua principal vantagem é a facilidade de cálculo.

É sempre interessante comparar as estimativas obtidas a partir de estimadores clássicos com estimativas robustas. Na literatura existem algumas propostas de estimadores robustos para os parâmetros da $G_{\xi, \mu, \sigma}$. Por exemplo, Dupuis, D. J. e Field, C. A. (1998) propõem um procedimento robusto, baseado em estimadores B-robustos e em funções de peso usadas para validar as observações. Este procedimento foi testado em dados reais e simulados, e sob contaminações ele se mostrou mais eficiente que os PWM, e bastante útil como diagnóstico para a qualidade do ajuste do modelo.

Veja exemplos de ajustes desses modelos em Mendes, B. V. M. (2004). Foram coletados máximos e mínimos mensais ($n = 22$) a partir dos log-retornos diários de algumas séries, e as distribuições foram ajustadas utilizando o método dos L-momentos. Para o conjunto de dados utilizados todas as estimativas de ξ foram positivas (Fréchet), caracterizando distribuições de retornos diários com caudas moderadamente pesadas. É interessante notar que as caudas tiveram pesos diferentes embora a distribuição dos retornos diários mostrasse ser simétrica.

Testes Estatísticos.

Qualquer que seja o método de estimação utilizado, é sempre recomendável testar-se formalmente as suposições do modelo e algumas hipóteses relativas aos parâmetros estimados. No caso da GEV, deve-se sempre testar a hipótese nula de $\xi = 0$, devido à maior simplicidade da expressão da f.d.a. Gumbel.

No caso de ajuste por L-momentos, o teste da qualidade do ajuste de Sherman (Sherman, B. (1957)) pode ser usado para testar o erro cometido pela substituição da distribuição exata (n finito) pela distribuição assintótica. Assim, o teste de Sherman tem hipótese nula H_0 : “Os extremos seguem uma f.d.a. GEV”. A estatística teste é

$$\frac{W_m - E_m}{D_m}$$

onde

$$W_m = \frac{1}{2} \sum_{i=1}^{m+1} |G_{\xi}(m_{(i)}) - G_{\xi}(m_{(i-1)}) - 1/(m+1)| ,$$

$$E_m = \left(\frac{m}{m+1} \right)^{m+1} ,$$

e

$$E_m^2 + D_m^2 = \frac{2m^{m+2} + m(m-1)^{m+2}}{(m+2)(m+1)^{m+2}} ,$$

onde G_{ξ} é a distribuição GEV com as estimativas obtidas, $G_{\xi}(m_{(0)}) = 0$, $G_{\xi}(m_{(m+1)}) = 1$, e onde a notação $m_{(k)}$ representa o máximo ordenado. A estatística teste é assintoticamente Normal e o p-valor deve ser calculado comparando-se com a cauda direita de uma Normal padrão. Por exemplo, para os mínimos do Brasil o p-valor foi de 0.422, portanto temos evidências estatísticas de que a cauda

esquerda do índice do Brasil é pesada, e os mínimos seguem uma distribuição Fréchet com índice de cauda $1/\xi = 4.28$.

Para se testar a hipótese de que os extremos seguem a distribuição mais simples Gumbel (um parâmetro a menos para se estimar), podemos usar o teste de Hosking dado em Hosking, J. R. M., Wallis, J. R. e Wood, E. F. (1985) cuja hipótese nula é $H_0 : \xi = 0$. Neste caso, a estatística teste é

$$Z = (-\hat{\xi}) \left(\frac{m}{0.5633} \right)^{\frac{1}{2}}$$

onde $\hat{\xi}$ é a estimativa por L-momentos. Esta estatística deve ser comparada com os valores de uma Normal padrão. Valores positivos significantes de Z rejeitam a hipótese nula em favor da Weibull, ao passo que valores negativos significantes de Z rejeitam a hipótese nula em favor da Fréchet. Por exemplo, este teste rejeita a hipótese nula em favor da Fréchet para os mínimos do índice do Brasil com p-valor de 0.0075.

No caso de estimação por máxima verossimilhança outros testes estão disponíveis. Para se testar a suposição de que os dados seguem a distribuição GEV pode-se usar as estatísticas de Kolmogorov-Smirnov D^+ , D^- , e D , e a estatística de Kuiper V dadas em Chandra, M., Singpurwalla, N. D. e Stephens, M. A. (1981), e definidas como

$$\begin{aligned} D^+ &= \max_i \left\{ \frac{i}{m} - G_\xi(m_{(i)}) \right\}, \\ D^- &= \max_i \left\{ G_\xi(m_{(i)}) - \frac{i-1}{m} \right\}, \\ D &= \max(D^+, D^-), \\ V &= D^+ + D^-, \end{aligned}$$

onde, novamente, $m_{(i)}$ são os máximos ordenados e G_ξ é a distribuição GEV com as estimativas obtidas.

Tabela 2.4: Valores críticos para as estatísticas teste D^+ , D^- , D , e V .

Nível de Significância	$\sqrt{m}D^+$	$\sqrt{m}D^-$	$\sqrt{m}D$	$\sqrt{m}V$
$m = 50$				
1%	0.940	0.944	0.988	1.639
5%	0.796	0.796	0.856	1.428
$m = \infty$				
1%	0.957	0.957	1.007	1.671
5%	0.808	0.808	0.874	1.477

Alguns valores críticos são dados na Tabela 2.4 para os níveis de significância de 1% e 5%, e para amostras de tamanho $m = 50$, e $m = \infty$ representando amostras grandes. Se os quatro valores calculados para $\sqrt{m}D^+$, $\sqrt{m}D^-$, $\sqrt{m}D$, e $\sqrt{m}V$ forem menores que os da tabela, *não* rejeitamos a H_0

de bom ajuste. No caso dos máximos mensais do Brasil, as estimativas por máxima verossimilhança são $\hat{\mu} = 3.351$, $\hat{\sigma} = 1.208$, e $\hat{\xi} = 0.382$. Os valores das estatísticas teste são $\sqrt{m}D^+ = 0.3385$, $\sqrt{m}D^- = 0.3069$, $\sqrt{m}D = 0.3385$, e $\sqrt{m}V = 0.6455$, todas elas menores que os valores críticos dados na Tabela 2.4 para o nível de significância de 5%. Portanto temos evidências estatísticas da boa qualidade do ajuste Fréchet para os máximos do Brasil.

Assim como no caso das estimativas por L-momentos, também para as estimativas por máxima verossimilhança pode-se testar se ξ é estatisticamente zero. Então, para testar a hipótese nula de que a distribuição dos extremos é Gumbel (modelo mais simples), usamos o teste da razão das verossimilhanças cuja estatística teste é

$$\Lambda = -2 (\mathcal{L}_{GUMBEL} - \mathcal{L}_{GEV}) ,$$

onde \mathcal{L}_{GUMBEL} e \mathcal{L}_{GEV} representam as log-verossimilhanças calculadas como em (2.21) usando respectivamente as densidades da Gumbel, $G_{\hat{\mu}, \hat{\sigma}}$, e da GEV, $G_{\hat{\xi}, \hat{\mu}, \hat{\sigma}}$, com as respectivas estimativas por máxima verossimilhança (veja [AP2.9]).

A estatística teste Λ deve ser comparada com uma χ_1^2 , para algum nível de significância preestabelecido. Seja $x_1^2(1-\alpha)$ o quantil de probabilidade $1-\alpha$ da χ_1^2 . Para $\alpha = 0.05$, temos que $x_1^2(0.95) = 3.8415$. Por exemplo, no caso dos máximos mensais do Brasil, a estatística teste Λ é igual a 14.8, e rejeitamos fortemente que os máximos mensais do Brasil sigam uma distribuição Gumbel.

Intervalos de confiança assintóticos para as estimativas $(\hat{\xi}, \hat{\mu}, \hat{\sigma})$ são construídos usando os resultados da teoria assintótica dos EMV. O S-Plus e o R fornecem esses intervalos. Alternativamente, podemos construir intervalos de confiança usando o método bootstrap (Efron, B. e Tibshirani, R. J. (1993)), que neste caso, não oferece nenhuma dificuldade de implementação.

Em toda análise estatística, para se acessar a qualidade do ajuste, além dos testes estatísticos formais, devem ser feitas algumas análises gráficas. Por exemplo, pode-se fazer: (i) a comparação da f.d.a. empírica com a f.d.a. ajustada, superpostas em um mesmo gráfico (veja exercício no final deste capítulo); (ii) a comparação da densidade não paramétrica e da densidade estimada, superpostas a um histograma; etc. A análise de resíduos é sempre recomendada. No caso Fréchet, quando $\hat{\xi} > 0$, os resíduos

$$w_i = \left(1 + \hat{\xi} \frac{(m_i - \hat{\mu})}{\hat{\sigma}} \right)^{-\frac{1}{\hat{\xi}}} , \quad i = 1, \dots, m , \quad (2.25)$$

seguem (veja Hosking, J. R. M. e Wallis, J. R. (1997)) a distribuição Exponencial(1). A Figura 2.7 mostra à esquerda a boa aderência da Fréchet estimada por L-momentos ao histograma dos mínimos da China. À direita, temos o QQ-plot dos resíduos (2.25) versus os quantís da distribuição de referência Exponencial(1). Finalmente, como em qualquer análise estatística que envolva escolhas, subjetivas ou não, análises de sensibilidade são recomendadas. Por exemplo, podemos verificar a estabilidade das estimativas de ξ , μ , e σ em relação ao tamanho n do bloco. O S-Plus tem já uma função para esta análise de sensibilidade, quando é feito um gráfico das estimativas dos parâmetros versus o tamanho do bloco (veja os exercícios no final deste capítulo).

Estimação do Índice Extremal.

De acordo com Leadbetter, M., Lindgren, G. e Rootzén, H. (1983), para séries estacionárias apresentando apenas fraca dependência de memória longa, pode-se ainda assumir que os máximos coletados

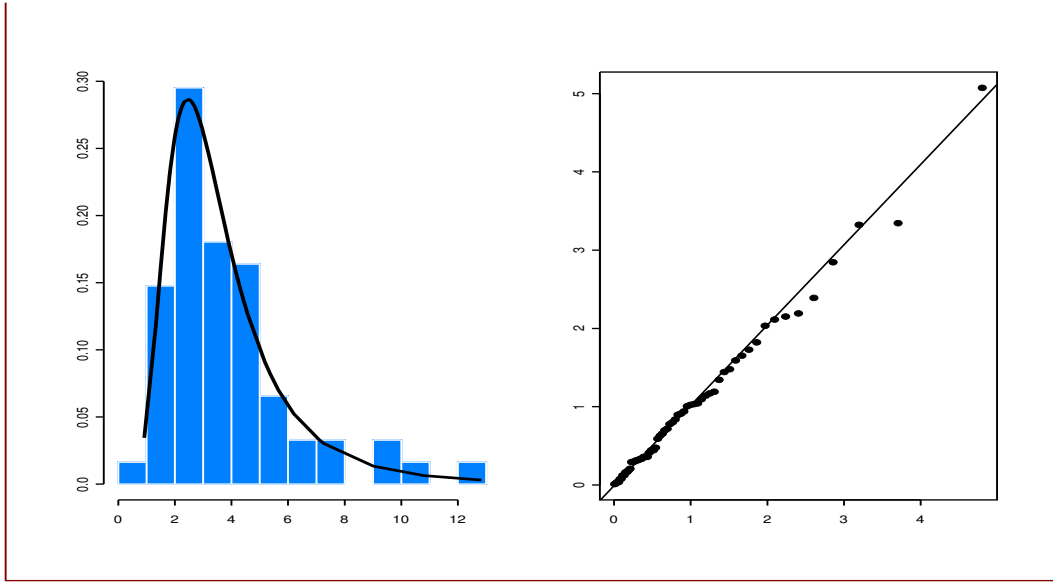


Figura 2.7: À esquerda a densidade Fréchet estimada e o histograma dos valores absolutos dos mínimos da China. À direita, o Q-Q-plot dos resíduos versus a distribuição de referência Exponencial(1).

em blocos de tamanho suficientemente grande possam ser bem modelados por uma distribuição GEV. Contudo, se esta série apresentar tendência a formar *clusters* de valores extremos, a convergência para a distribuição limite será mais lenta que no caso i.i.d. A fórmula (2.20) introduziu o índice extremal e forneceu a correção para o caso de existir dependência na cauda extrema de F .

Um estimador simples (veja em Hsing, T., Hüsler, J. e Leadbetter, M.R. (1988)) para θ é a razão entre o número K_u de blocos de tamanho N onde o máximo excedeu um limiar especificado u , e o número T_u de observações que excederam u durante todo o período sob análise. Temos

$$\hat{\theta} = \frac{K_u}{T_u}.$$

É claro que a estimativa de θ dependerá das escolhas N e u . O tamanho do bloco N não pode ser muito grande e não precisa necessariamente ser o mesmo do máximo sob estudo. Pode-se especificar u a partir de um certo percentual de observações nas caudas, 5% a 3%. Uma outra possibilidade é observar graficamente a sensibilidade da estimativa do índice extremal em relação à escolha do limiar e à escolha do tamanho do bloco.

Na literatura existe outra proposta para a estimação de θ com melhores propriedades estatísticas (Smith, R. L. e Weissman, I. (1994), Embrechts, P., Klüppelberg, C. e Mikosch, T. (1997)) e assintoticamente equivalente ao estimador acima. É

$$\hat{\theta} = \frac{1}{m} \frac{\log \left(1 - \frac{K_u}{m} \right)}{\log \left(1 - \frac{T_u}{T} \right)} \quad (2.26)$$

onde T_u , T , K_u e m já foram definidos. Como antes, o limiar u pode ser escolhido tal que $\frac{T_u}{T} * 100\%$

represente um certo percentual dos dados. A Figura 2.8 mostra o índice extremal para as perdas do Brasil em função do tamanho do bloco e do valor do limiar, e estimado a partir de (2.26). A partir desta figura poderia se escolher $N = 5$ e $\theta = 0.82$. Esta estimativa é confirmada quando se fixa o limiar a partir da escolha de 5% de observações na cauda.

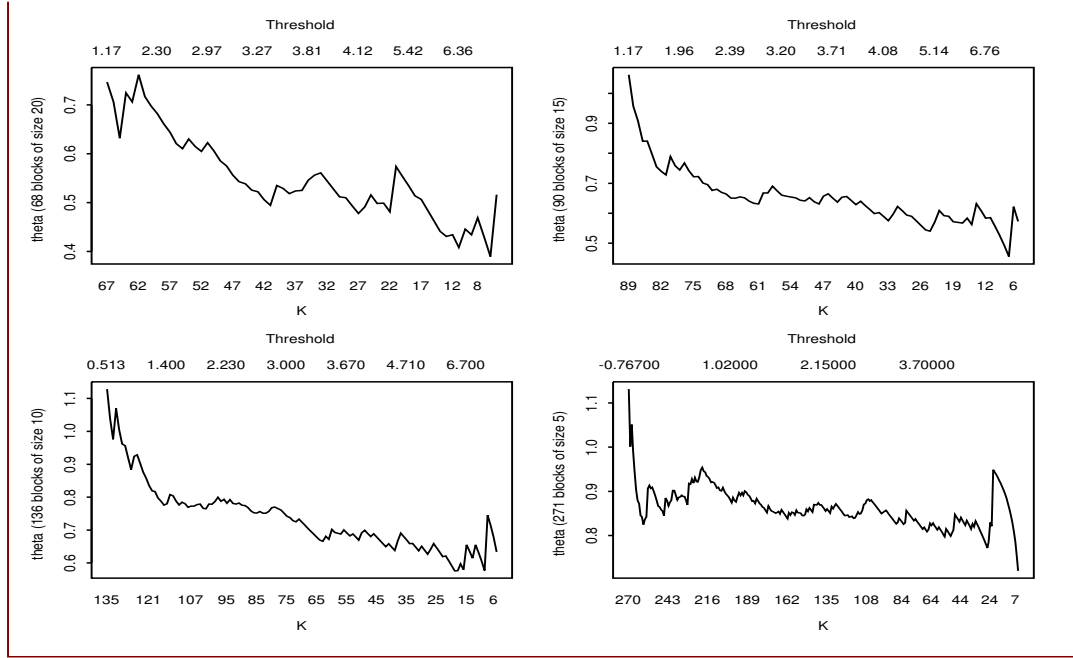


Figura 2.8: Estimativas do índice extremal para várias opções de tamanho N de bloco e limiares, e para a cauda esquerda do índice do Brasil.

Podemos interpretar $n\theta$ como o número de “conglomerados independentes” em n observações i.i.d. Assim, durante o mês de crise, o número médio de observações agrupadas é $n/n\theta = 1/\theta$, o tamanho médio dos conglomerados. O índice extremal é uma ferramenta interessante para discriminar as caudas de uma distribuição. Por exemplo, calculado para uma série de retornos, ele indica se retornos extremos se aglomeram mais frequentemente em uma das caudas (veja Jondeau, E. e Rockinger, M. (2003) e exercício no final deste capítulo).

Cenários de Stress.

A modelagem através da distribuição GEV permite a simulação e análise de eventos extremos e adversos. Muitas vezes, quer-se prevenir de algum evento extremo, e a melhor maneira de definir esse evento é atribuir a ele uma periodicidade. Em geral fixa-se um tempo médio de espera bastante grande para uma nova ocorrência, como por exemplo 100 meses, para então se obter seu valor. Esta técnica foi utilizada no exemplo sobre um cenário adverso para o Brasil no ano de 2008.

O evento de t -meses (Zangari, P., 1997) ou, mais geralmente, evento de t -períodos, é um conceito muito simples, sendo basicamente a associação de um quantil da distribuição dos máximos com o

número médio t de períodos (blocos de tamanho n) entre suas ocorrências. O evento de t -meses, denotado por E_t satisfaz

$$Pr\{M_n > E_t\} = \frac{1}{t}, \quad (2.27)$$

onde $t > 1$, sendo portanto um evento da distribuição do máximo (ou valor absoluto do mínimo) cujo tempo de espera médio é de t meses. Notemos que $\frac{1}{t} = 1 - Pr\{M_n \leq E_t\} = 1 - G(E_t)$. Assim, se $t = 50$ meses, $1 - \frac{1}{50} = 0.98$, e então E_{50} é o 0.98-quantil da distribuição do máximo mensal.

Tendo estimado os parâmetros ξ , μ , e σ , podemos calcular o evento de t -meses usando a inversa da f.d.a. GEV. Isto é:

$$E_t = G_{\hat{\xi}, \hat{\mu}, \hat{\sigma}}^{-1}\left(1 - \frac{1}{t}\right) = \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}} \left(1 - \left(-\log\left(1 - \frac{1}{t}\right)\right)^{-\hat{\xi}}\right), \quad (2.28)$$

onde $G_{\hat{\xi}, \hat{\mu}, \hat{\sigma}}^{-1}(\cdot)$ representa a inversa da f.d.a. $G_{\xi, \sigma, \mu}(\cdot)$ com as estimativas obtidas.

É esclarecedor expressar o evento de t -meses a partir dos quantís (extremos) da distribuição subjacente F , já que $F(x) = (G_{\xi, \mu, \sigma}(x))^{1/n}$. Por exemplo, no caso de uma série de retornos diários e extremos coletados em blocos mensais, o evento de 50-meses é o 0.999082-quantil desta distribuição subjacente. Assim, a probabilidade de que sejam observados retornos diários mais extremos que o E_{50} é de 0.000918. Uma maior acurácia podemos ter se levarmos em consideração a heteroscedasticidade condicional dos dados, isto é, a existência dos clusters de volatilidade, bastando para isto incorporar a estimativa de θ . Notemos aqui o papel e vantagem do uso da TVE: teria sido bastante difícil estimar a partir da F e com precisão um evento tão raro.

Os valores de eventos de 6, 12, e 48 meses correspondem a quantis da F de probabilidades de excedência iguais a 0.00825, 0.00395, e 0.00096, respectivamente. Sugerimos, como exercício, o cálculo desses cenários de stress baseados em eventos de 6, 12, e 48 meses.

Uma outra maneira de se construir cenários de stress como acima, é estabelecendo um certo valor extremo, e obter o tempo de espera até a ocorrência de retornos além deste limiar. Alternativamente, pode-se fixar um tempo de espera, e calcular-se a probabilidade de observar um retorno além de um limiar prefixado. É fácil ver a grande aplicabilidade desta abordagem em finanças. Essas aplicações encontram-se em Mendes, B. V. M. (2004).

Por exemplo, a Figura 2.9 mostra no eixo horizontal o tempo médio de espera, e no eixo vertical o valor u do limiar para os mínimos mensais do Brasil e da China. Observamos grande diferença de escala. Por exemplo, o tempo médio de espera de (aproximadamente) 60 meses corresponde à ultrapassagem do limiar $u = -15$ para o Brasil, ao passo que corresponde a $u = -11$ para a China.

Modelo para as r -maiores estatísticas de ordem

Uma crítica imputada ao modelo GEV é que ele despreza algumas observações extremas, utilizando apenas o máximo de cada bloco. A análise estatística é então realizada usando-se poucos dados, e com isto as estimativas dos parâmetros e de outras funções de interesse apresentam grande variabilidade. Pode-se estender os resultados anteriores para um modelo que utiliza as r -maiores estatísticas de ordem dentro de cada bloco, r pequeno, veja Smith, R. L. (1986), Weissman, I. (1978), Tawn, J.

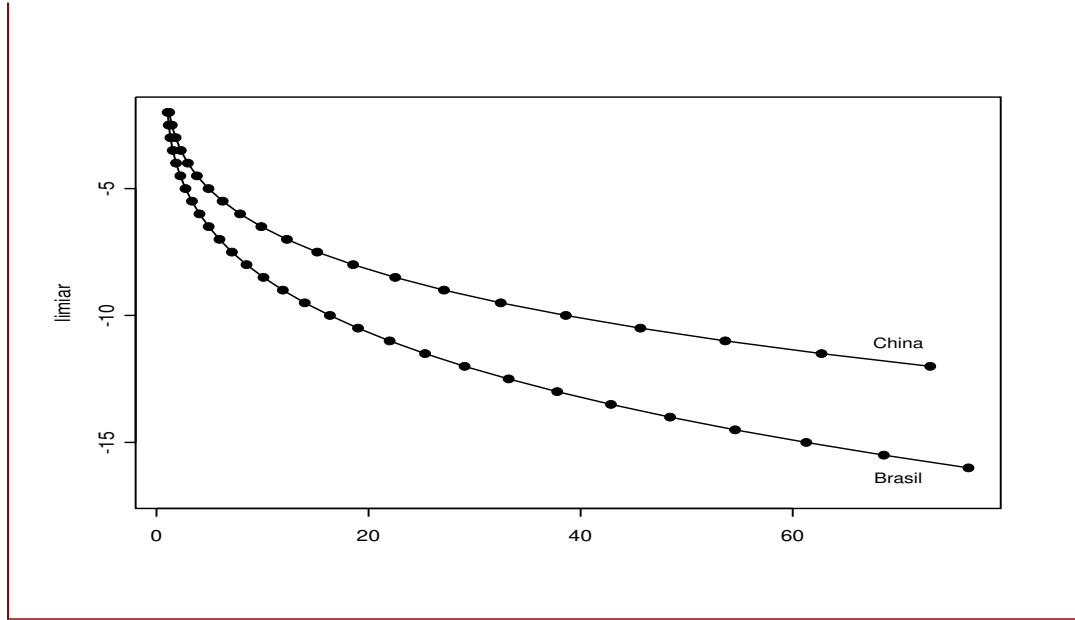


Figura 2.9: O tempo médio de espera versus o limiar u para os mínimos mensais do Brasil e da China.

(1988b). Através da inclusão de dados adicionais relevantes, este procedimento leva à uma maior precisão das estimativas, quando comparado à abordagem clássica dada pela GEV.

Considere novamente $\{r_t\}$, $t = 1, 2, \dots, n$, uma sequência de v.a.'s i.i.d. com distribuição F , e seja $M_n^{(k)}$ a k -ésima maior estatística de ordem entre as n variáveis, $k = 1, 2, \dots, r$. Assumindo as premissas do teorema de Fisher e Tippet para as outras estatísticas de ordens extremas, temos que para k fixo e quando $n \rightarrow \infty$

$$P\left(\frac{M_n^{(k)} - b_n}{a_n} \leq z\right) \rightarrow G_k(z)$$

em $\{z : 1 + \xi(z - \mu)/\sigma > 0\}$, onde

$$G_k(z) = \exp\{-\tau(z)\} \sum_{s=0}^{k-1} \frac{\tau(z)^s}{s!}, \quad (2.29)$$

com

$$\tau(z) = \left[1 + \xi \left(\frac{z - \mu}{\sigma}\right)\right]^{-1/\xi}.$$

O resultado acima quer dizer que a k -ésima maior estatística de ordem em um bloco de tamanho n é padronizada exatamente da mesma forma que o máximo $M_n^{(1)} = M_n$. Os parâmetros da distribuição limite são os mesmos da GEV. Novamente, em termos de estimação, as constantes normalizadoras desconhecidas são absorvidas pelos parâmetros de locação e escala, já que a estimação é feita para um dos membros da família.

Seja $\mathbf{M}^{(r)} = (M_n^{(1)}, \dots, M_n^{(r)})$, o vetor das r -maiores estatísticas de ordem para n fixo e para algum valor de r . É bastante intuitivo que as componentes $M_n^{(k)}$ não sejam independentes. Precisamos por isto do modelo conjunto para as r -maiores. Para r fixo, a distribuição limite, quando $n \rightarrow \infty$, de

$$\tilde{\mathbf{M}}_n^{(r)} = \left(\frac{M_n^{(1)} - b_n}{a_n}, \dots, \frac{M_n^{(r)} - b_n}{a_n} \right)$$

pertence à família que tem função densidade de probabilidade conjunta dada por

$$f(z^{(1)}, \dots, z^{(r)}) = \exp \left\{ - \left[1 + \xi \left(\frac{z^{(r)} - \mu}{\sigma} \right) \right]^{-1/\xi} \right\} \times \prod_{k=1}^r \sigma^{-1} \left[1 + \xi \left(\frac{z^{(k)} - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi} - 1} \quad (2.30)$$

onde $-\infty < \mu < \infty$, $\sigma > 0$ e $-\infty < \xi < \infty$; $z^{(r)} \leq z^{(r-1)} \leq \dots \leq z^{(1)}$; e $z^{(k)} : 1 + \xi(z^{(k)} - \mu)/\sigma > 0$ para $k = 1, \dots, r$.

Para o caso $r = 1$, (2.30) fica reduzida à família GEV de funções de densidade. O caso $\xi = 0$ em (2.30) é interpretado como a forma limite quando $\xi \rightarrow 0$, e conduz à família das funções de densidade

$$f(z^{(1)}, \dots, z^{(r)}) = \exp \left\{ - \exp \left[- \left(\frac{z^{(r)} - \mu}{\sigma} \right) \right] \right\} \times \prod_{k=1}^r \sigma^{-1} \exp \left[- \left(\frac{z^{(k)} - \mu}{\sigma} \right) \right], \quad (2.31)$$

na qual o caso $r = 1$ fica reduzido à família da densidade Gumbel.

A estimação em geral é por máxima verossimilhança, veja exemplos em Mendes, B. V. M. (2004).

Modelo GPD para Excessos Além de um Limiar

Também com o intuito de responder às críticas quanto ao desperdício de observações extremas, há ainda outro modelo que pode ser utilizado para modelar as caudas das distribuições: a distribuição Generalizada de Pareto (GPD), que modela excessos além de um limiar alto. Os argumentos teóricos que conduzem a essa distribuição são atribuídos à Pickands, J. III (1975a).

Considere o retorno r_t ou, por simplicidade de notação r (já que a característica temporal da série agora não será levada em consideração), um limiar u , o excesso $Y = r - u$ e sua função de distribuição condicional F_u , $F_u(y) = P(Y \leq y | r > u)$, $y \geq 0$. A função de sobrevivência condicional do excesso Y é dada por

$$\bar{F}_u(y) = P(r - u > y | r > u) = \frac{\bar{F}(u + y)}{\bar{F}(u)}. \quad (2.32)$$

Segue então que

$$\bar{F}(u + y) = \bar{F}_u(y) \bar{F}(u), \quad (2.33)$$

expressão muito útil para o cálculo de medidas de risco no Capítulo 4.

A adequacidade da GPD para modelar os excessos além de um limiar alto se justifica a partir do fato de que a distribuição F de r pertencer ao domínio de atração da GEV. De acordo com de Haan, L. (1984), $F \in MDA(G_\xi)$ se e somente se existe uma função $a(\cdot)$ positiva e mensurável tal que para $1 + \xi x > 0$ tem-se

$$\lim_{u \uparrow} \frac{\bar{F}(u + xa(u))}{\bar{F}(u)} = \begin{cases} (1 + \xi x)^{-1/\xi}, & \text{se } \xi \neq 0 \\ e^{-x}, & \text{se } \xi = 0. \end{cases} \quad (2.34)$$

O resultado acima pode ser reescrito como: $F \in MDA(G_\xi)$ se e somente se

$$\lim_{u \uparrow} Pr\left\{\frac{X-u}{a(u)} > x | X > u\right\} = \begin{cases} (1 + \xi x)^{-1/\xi}, & \text{se } \xi \neq 0 \\ e^{-x}, & \text{se } \xi = 0. \end{cases} \quad (2.35)$$

Isto é, a expressão (2.35) nos dá uma aproximação para a distribuição dos excessos (apropriadamente reduzidos por um fator de escala) acima do limiar u . Notemos que o fator de escala apropriado é $a(u)$. De fato, a distribuição de Pareto generalizada padrão, denotada por P_ξ , tem f.d.a. dada por

$$P_\xi(y) = \begin{cases} 1 - (1 + \xi y)^{-1/\xi}, & \text{se } \xi \neq 0 \\ 1 - e^{-y}, & \text{se } \xi = 0 \end{cases} \quad (2.36)$$

onde $y \geq 0$ se $\xi \geq 0$, e $0 \leq y \leq -1/\xi$ se $\xi < 0$, e cuja densidade é

$$p_\xi(y) = (1 + \xi y)^{-(1+1/\xi)} \quad \text{para} \quad \begin{cases} 0 \leq y, & \text{se } \xi \geq 0 \\ 0 \leq y \leq \frac{1}{|\xi|}, & \text{se } \xi < 0. \end{cases} \quad (2.37)$$

A família de locação e escala $P_{\xi, \delta, \psi}$ pode ser obtida substituindo-se o argumento y por $(y - \delta)/\psi$, para $\delta \in \mathfrak{R}$, $\psi > 0$. Neste caso, os suportes $D(\xi, \delta, \psi)$ da distribuição são: $D(\xi, \delta, \psi) = [\delta, \infty]$ se $\xi \geq 0$, e $D(\xi, \delta, \psi) = [\delta, \delta - \psi/\xi]$ se $\xi < 0$. Veja no Apêndice [AP3.1] deste capítulo as expressões dos momentos da GPD, e as expressões das densidades com os três parâmetros. Dentro da classe $P_{\xi, \delta, \psi}$ a distribuição de maior interesse nas aplicações práticas em finanças é a $P_{\xi, 0, \psi}$ com $\xi \geq 0$ e y com suporte $D(\xi, 0, \psi)$.

Existe uma relação analítica simples entre a GPD e a GEV, dada por

$$P_\xi(x) = 1 + \ln(G_\xi(x)), \quad \ln(G_\xi(x)) > -1.$$

Isto explica por que as densidades da GPD possuem cauda extrema assintoticamente equivalentes às de uma GEV. A Figura 2.10 ilustra este fato e mostra a proximidade das caudas de algumas distribuições GPD com algumas GEV.

Assim como no caso da GEV, podemos obter três submodelos de distribuições padrões ($\delta = 0, \psi = 1$) que são os tipos

Tipo I. Exponencial, $\xi = 0$, $P_I(y) = 1 - e^{-y}$, $y \geq 0$.

Tipo II. Pareto, $\xi > 0$, $P_{II, \xi}(y) = 1 - y^{-\frac{1}{\xi}}$, $y \geq 1$.

Tipo III. Beta, $\xi < 0$, $P_{III, \xi}(y) = 1 - (-y)^{-\frac{1}{\xi}}$, $-1 \leq y \leq 0$.

Novamente, P_I pode ser interpretada como o limite da P_ξ quando $\xi \rightarrow 0$. Um caso particular bastante conhecido é o $P_{III, -1}$, a distribuição Uniforme $[-1, 0]$. A Figura 2.11 compara a densidade Exponencial com os outros dois tipos. Quando $\xi > 0$ têm-se caudas mais pesadas.

Se $F \in MDA(G_\xi)$ então F pertence ao domínio de atração POT (*Peaks Over Threshold*) de P_ξ . As distribuições GPD são as únicas POT-estáveis. Por exemplo, se $X \sim \text{Exponencial}(\lambda)$, a distribuição do excesso além do limiar u , $F_u(\cdot)$ é também Exponencial(λ). Isto é:

$$Pr\{Y > y | X > u\} = \frac{Pr\{X > y + u, X > u\}}{Pr\{X > u\}}$$

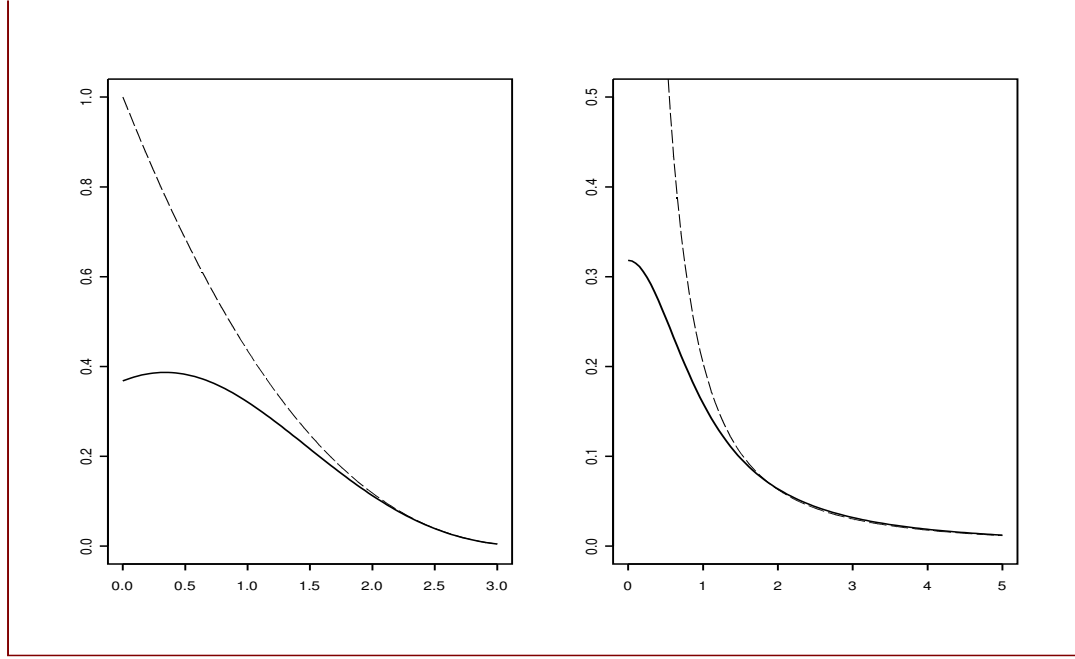


Figura 2.10: (Esquerda) Densidades: GPD (Beta), tracejada e GEV (Weibull), contínua, ambas com parâmetro $\xi = -0,3$. (Direita) Densidades: GPD (Pareto), tracejada com parâmetros $\xi = 1$, $\mu = -0,25$ e $\sigma = 1/\pi$ e Cauchy padrão, contínua.

$$= \frac{\exp\{-\lambda(u+y)\}}{\exp\{-\lambda u\}} = \exp\{-\lambda y\}.$$

Um outro resultado importante que garante o uso da GPD como uma distribuição adequada para F_u é o teorema de Pickands, J. III (1975a), que afirma que

$$\inf_{\xi} \lim_{u \uparrow w_F} \inf_{\sigma} \sup_{x > 0} |F_u(x) - P_{\xi, \psi(u)}(x)| = 0$$

se e somente se F está no domínio de atração de uma das três distribuições padrão de valores extremos. Aqui $w_F = \sup\{x : F_u(x) < 1\}$, o limite superior de F_u que pode ser finito ou infinito. Em outras palavras, existe um ξ que não depende de u , e um ψ que depende de u , tais que F_u se aproxima de $P_{\xi, \psi}$ sempre que u estiver suficientemente próximo de w_F .

Uma propriedade de grande interesse em ciências atuariais, em particular em seguros e resseguros, é a seguinte, dada em Embrechts, P., Klüppelberg, C. e Mikosch, T. (1997). Suponha que $y_i \in D(\xi, \psi)$, $i = 1, 2$. Então,

$$\frac{\bar{P}_{\xi, \psi}(y_1 + y_2)}{\bar{P}_{\xi, \psi}(y_1)} = \bar{P}_{\xi, \psi + \xi y_1}(y_2). \quad (2.38)$$

Isto significa que a classe das distribuições GPD é fechada em relação a mudanças no limiar. De fato, observe que o lado esquerdo de (2.38) é a probabilidade

$$Pr\{Y > y_1 + y_2 | Y > y_1\},$$

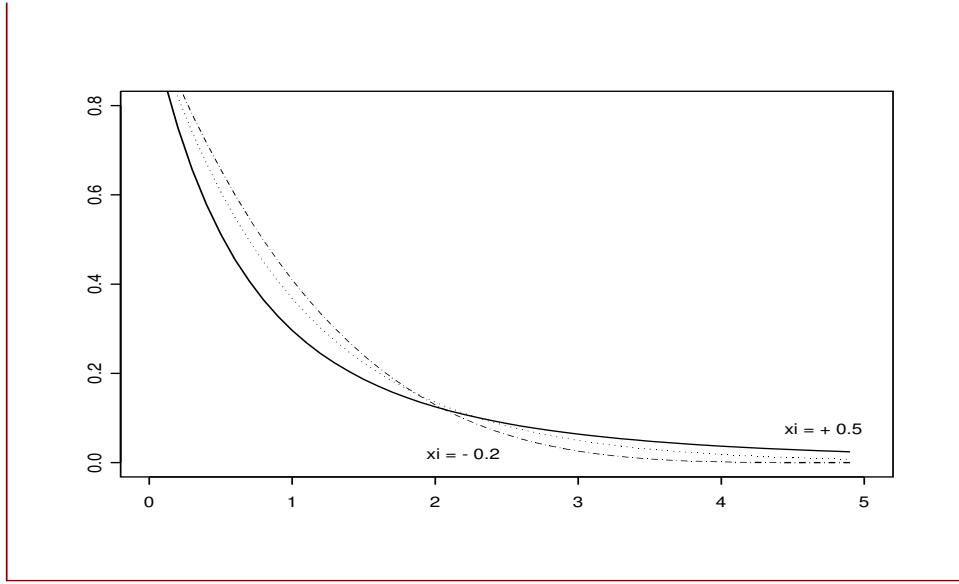


Figura 2.11: Densidades: Exponencial (pontilhada); Pareto $\xi = 0,5$ (contínua); Beta $\xi = -0.2$ (tracejada). As três densidades possuem locação zero e escala um.

isto é, a probabilidade de que a variável exceda o limiar $y_1 + y_2$ dado que excede y_1 . O lado direito de (2.38) diz que esta probabilidade é novamente uma GPD. Esta propriedade de fechamento é importante em resseguros, onde a perda em excesso é segurada. Notemos que após 11 de setembro de 2001, este tópico se tornou de grande interesse para as firmas de resseguros do mundo inteiro.

Funções Média e Mediana dos Excessos. A esperança da distribuição condicional dos excessos além de um limiar predefinido u , $u < x^+$, ou a média dos excessos, é uma função de u . Podemos então definir a função Média dos Excessos, $e(u)$, como

$$e(u) = E[r - u | r > u].$$

Pode-se mostrar que para toda f.d.a. contínua tem-se

$$\bar{F}(x) = \frac{e(0)}{e(x)} \exp \left\{ - \int_0^x \frac{1}{e(u)} du \right\}, \quad x > 0,$$

de onde concluímos que toda F contínua é unicamente determinada por sua $e(u)$. Vejamos alguns exemplos. No caso de $Y \sim P_{\xi, \psi}$ com $\xi < 1$, então, para $u < x^+$,

$$e(u) = \frac{\psi + \xi u}{1 - \xi}, \quad \psi + u\xi > 0.$$

Notemos que a $e(u)$ de uma GPD é função linear de u . Se $\xi > 0$, caso particular Pareto, a reta tem inclinação positiva. Se $\xi < 0$, a inclinação de $e(u)$ é negativa. Se Y tem distribuição Exponencial(λ),

então a f.d.a de Y é $e^{-\lambda x}$, $\lambda > 0$ e

$$e(u) = \frac{1}{\lambda},$$

e neste caso $e(u)$ não depende do limiar.

Uma ferramenta gráfica exploratória é o gráfico da função Média dos Excessos empírica, $e_{N_u}(u)$, ou de sua versão robusta, a função Mediana dos Excessos, $e_{N_u}^*(u)$, *versus* o limiar u . Veja também outra versão robusta dada pela Média Podada dos Excessos no Apêndice [AP2.8]. Dada uma amostra r_1, \dots, r_T , essas funções são definidas como

$$e_{N_u}(u) = \frac{1}{N_u} \sum_i^{N_u} Y_i(u)$$

$$e_{N_u}^*(u) = \text{Mediana}\{Y_i(u), i = 1, \dots, N_u\}$$

onde $Y_i(u)$ são os excessos além do limiar u , isto é, os N_u valores $r_i - u > 0$.

Escolha do Limiar. Temos aqui um caso típico de *trade-off* entre vício e variância. Na escolha do limiar ótimo, um valor para u muito “alto” implicará em um número pequeno de observações na cauda, podendo resultar em uma maior variabilidade dos estimadores. Por outro lado, um limiar não suficientemente alto não satisfaz às suposições teóricas e pode resultar em estimativas distorcidas.

Vários autores têm investigado o problema da escolha do limiar u . Pickands, J. III (1975a) sugere um procedimento simultâneo para a escolha do limiar e estimação dos parâmetros ξ e ψ da GPD (veja também Smith, R. L. (1987)). Pickands, J. III (1975b) propôs um método para a escolha das k estatísticas de ordem definindo o limiar ótimo. Outros trabalhos, como Mendes, B. V. M. e Lopes, H. F. (2004) e Mendes, B. V. M. (2005a), não abordam explicitamente o problema da escolha do limiar, mas o obtêm como subproduto após o ajuste de um modelo de misturas.

Uma outra possibilidade é a de se usar as funções Média e Mediana dos Excessos empíricas. A técnica gráfica exploratória baseada nessas funções podem nos auxiliar na tarefa da escolha do limiar u alto suficiente para que a aproximação da distribuição dos excessos por uma GPD seja justificada. Assim, devemos procurar por aquele(s) valor(es) de u tais que a partir dele a $e_{N_u}(u)$ (ou a $e_{N_u}^*(u)$) se torna aproximadamente linear.

Estimação da GPD. Vários procedimentos para estimar os parâmetros ξ e ψ da GPD têm sido propostos na literatura. Por exemplo, um método sugerido em Pickands, J. III (1975a) também estima o limiar. Pickands, J. III (1984) propôs uma abordagem não paramétrica. Hosking, J. R. M. e Wallis, J. R. (1987) consideram o método dos estimadores PWM. Smith, R. L. (1987) propôs um estimador baseado nos estimadores de máxima verossimilhança que seria altamente eficiente se o limiar escolhido fosse ótimo. Pickands, J. III (1975b) considerou estimadores de Bayes usando três estatísticas de ordem, inclusive o máximo.

Considere os dados de excessos y_1, y_2, \dots, y_{N_u} . A densidade da GPD com locação zero e escala 1 é:

$$p_{\xi, \psi}(y) = \frac{1}{\psi} \left(1 + \xi \frac{y}{\psi} \right)^{-(1+1/\xi)} \quad \text{se } y \in D(\xi, \psi). \quad (2.39)$$

Os EMV maximizam a função de log-verossimilhança

$$\mathcal{L}((\xi, \psi, u); y_1, \dots, y_{N_u}) = -N_u \log(\psi) - \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^{N_u} \ln \left(1 + \frac{\xi}{\psi} y_i\right), \quad (2.40)$$

onde $y_i \in D(\xi, \psi)$ (veja propriedades desses estimadores em Smith, R. L. (1987)). Recordemos que $D(\xi, \psi) = [0, \infty)$ se $\xi \geq 0$. Podemos agora derivar (2.40) em relação aos parâmetros e obter as equações de log-verossimilhança, as quais podem ser resolvidas numericamente. Os estimadores $(\hat{\xi}, \hat{\psi})$ possuem as boas propriedades dos EMV quando $\xi > -\frac{1}{2}$. Neste caso eles são conjuntamente assintoticamente normais, isto é:

$$N_u^{1/2} \left(\hat{\xi} - \xi, \frac{\hat{\psi}}{\psi} - 1 \right) \xrightarrow{d} N_2(0, M^{-1}), \quad N_u \rightarrow \infty,$$

onde

$$M^{-1} = (1 + \xi) \begin{pmatrix} 1 + \xi & 1 \\ 1 & 2 \end{pmatrix},$$

e onde N_2 denota a distribuição normal bivariada.

Os estimadores de L-momentos para a GPD (com parâmetro de locação δ) podem ser obtidos a partir das equações

$$\lambda_1 = \delta + \frac{\psi}{1-\xi}$$

$$\lambda_2 = \frac{\psi}{(1-\xi)(2-\xi)}$$

$$\frac{\lambda_3}{\lambda_2} = \tau_3 = \frac{(1+\xi)}{(3-\xi)}$$

$$\frac{\lambda_4}{\lambda_2} = \tau_4 = \frac{(1+\xi)(2+\xi)}{(3-\xi)(4-\xi)}$$

onde a quantidade τ_k , versão adimensional dos L-momentos, é

$$\tau_k = \frac{\lambda_k}{\lambda_2}, k = 3, 4, \dots,$$

e onde a relação entre τ_3 e τ_4 é dada por

$$\tau_4 = \frac{\tau_3(1 + 5\tau_3)}{(5 + \tau_3)}.$$

Para estimar os parâmetros ψ e ξ , substituímos nas equações acima os L-momentos populacionais pelos seus correspondentes amostrais l_1 , l_2 e l_3 . No nosso caso o parâmetro de locação δ é conhecido e é o próprio limiar u . Assim, os outros dois parâmetros ψ e ξ são dados por:

$$\hat{\xi} = \frac{u - l_1}{l_2} + 2$$

$$\hat{\psi} = (1 - \hat{\xi})(l_1 - u).$$

Novamente, suas principais vantagens (Hosking, J. R. M. e Wallis, J. R (1997)) são unicidade, vício e variância pequenos, rapidez de cálculo, e robustez quando comparados aos EMV.

O estimador de máxima verossimilhança para ξ no caso particular quando P_ξ é Pareto, isto é, quando $\xi > 0$, é o estimador de Hill (Hill, B. (1975)) (veja exercício 10 no final deste capítulo). Considere a amostra r_1, \dots, r_T e as estatísticas de ordem $r_{(1)} \leq r_{(2)} \leq \dots \leq r_{(T)}$. O estimador de Hill é dado por

$$\hat{\xi}_H = \frac{1}{k} \sum_{i=1}^k [\ln(r_{(T+1-i)}) - \ln(r_{(T-k)})] \quad (2.41)$$

onde $r_{(T-k)}$ é o limiar.

Como os dados r_1, \dots, r_T possuem distribuição comum F e não P_ξ as boas propriedades usuais dos EMV não se aplicam ao $\hat{\xi}_H$ que utiliza apenas uma fração k/T da amostra. Claro que quanto maior T maior deve ser $k = k(T)$. Assintoticamente devemos ter

$$k(T) \rightarrow \infty, \quad k(T)/T \rightarrow 0, \quad \text{quando } T \rightarrow \infty.$$

Se $F \in MDA(G)$, o estimador de Hill é assintoticamente não viciado mas sofre do problema da escolha do limiar (ou de k) e do *trade off* vício-variância. Tem-se que $\sqrt{k}(\hat{\xi}_H - \xi)$ é assintoticamente normal com variância ξ^2 .

A estimação de ξ é importante devido ao fato de ser $1/\xi$ o índice de cauda da distribuição F , o qual, além de medir o peso da cauda, também determina o número de momentos finitos de F , o que tem motivado a pesquisa sobre o estimador de Hill. Jansen, D. W. e de Vries, C. G. (1991) estudam o estimador dado em (2.41) e provam que este estimador é consistente até mesmo no caso das v.a.'s não serem independentes, e propõem um procedimento para selecionar k para qualquer tamanho de amostra T e qualquer F .

Em Dacorogna, M., Müller, U., Pictel, O. e de Vries, C. (1995), os autores derivam uma aproximação assintótica para o vício do estimador de Hill dentro de uma certa classe de distribuições. Para esta mesma classe de distribuições Hall, P. (1990) obtém a variância assintótica do $\hat{\xi}_H$ mostrando que ela depende de k ($(1/k)\xi^2$), novamente colocando o *trade-off*, já que quanto menor k menor o vício. A partir desses resultados, Huisman, R., Koedijk, K. G., Kool, C. J. M. e Palm, F. (2001) propõem um procedimento para corrigir o vício do estimador de Hill em amostras pequenas, o qual elimina o problema da escolha do limiar.

A proposta de Huisman, R., Koedijk, K. G., Kool, C. J. M. e Palm, F. (2001) se baseia numa simplificação da fórmula para o vício dada em Dacorogna, M., Müller, U., Pictel, O. e de Vries, C. (1995). Eles argumentam que até um certo valor para k , digamos, K , o estimador de Hill parece crescer linearmente com k . Em resumo, eles sugerem que

$$\hat{\xi}_H = \beta_0 + \beta_1 k + \epsilon(k), \quad k = 1, \dots, K.$$

Os autores sugerem estimar por mínimos quadrados (MQ) o estimador de Hill (2.41) para $k = 1, \dots, K$ e obter os coeficientes da regressão β_0 e β_1 . Como o estimador de Hill se torna não viciado quando k se aproxima de zero, justifica-se definir como o estimador final corrigido de Hill a estimativa de β_0 . Para a estimação dos parâmetros da regressão os autores mostram que a escolha de K não é tão

crucial. Huisman, R., Koedijk, K. G., Kool, C. J. M. e Palm, F. (2001) sugerem uma melhoria neste método utilizando mínimos quadrados ponderados (MQP), já que a variância do estimador de Hill não é constante. Além disto, os “dados” para a estimação da regressão não são independentes, já que existem dados comuns para os vários k 's. O peso seria então o recíproco da raiz quadrada de k .

Podemos adicionar um pouco de robustez a este processo e estimar robustamente a regressão linear acima. Vários estimadores robustos estão disponíveis no SPlus e no R. Rodamos alguns experimentos de simulação e verificamos que, ao contrário do que ocorre com as estimativas obtidas a partir dos MQ e dos MQP, a estimativa robusta *não* depende do valor de K . Por exemplo, utilizando a distribuição t com 4 g.l., os estimadores robustos de regressão MM-estimadores de Yohai, V., Stahel, W.A., e Zamar, R.H. (1991), e fixando $K = 450$ (veja Figura 2.12), a média de 10 mil simulações resultou nos valores para o índice de cauda $1/\xi$, cujo verdadeiro valor é 4, de 7.0085, 4.3824, 5.1533, respectivamente para a solução MQ, MM, MQP, mostrando que apenas o estimador robusto não depende de K . Quando fixamos $K = 300$, como sugere a Figura 2.12, obtivemos os melhores resultados de 4.2634, 4.3698, 4.2156, respectivamente para os métodos MQ, MM, e MQP, ressaltando a excelente estabilidade da nossa proposta robusta. Outros experimentos de simulação baseados em outras F 's podem ser facilmente rodados.

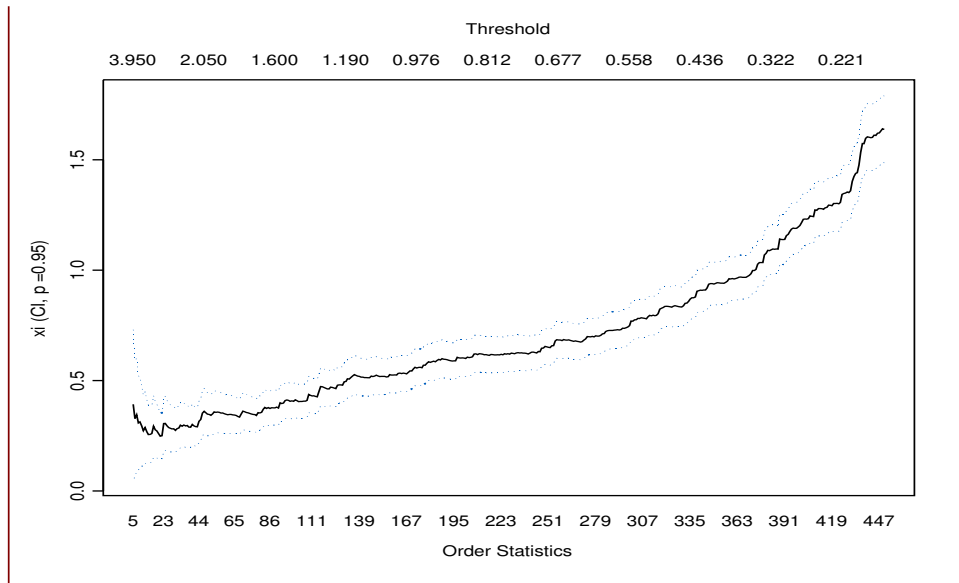


Figura 2.12: Gráfico de k versus estimativas do estimador de Hill correspondente.

Entretanto, para a maioria dos estimadores existe a questão da sensibilidade das estimativas quanto à escolha do limiar. A Figura 2.13 exibe a sensibilidade dos EMV de ξ em relação à escolha do limiar u , para a cauda esquerda do índice Brasil. A figura mostra a estimativa pontual juntamente com o intervalo de 95% de confiança. No eixo horizontal temos o número de observações usadas no processo de estimação, variando de 300 a 25, e no alto os limiares correspondentes. Notemos que 20% dos dados de retornos corresponde a 270 excessos, um número bem razoável para definir a cauda. Neste caso

não tivemos muita dificuldade para definir a estimativa de ξ próxima de 0.10, mas em geral iremos precisar de outras ferramentas para decidir sobre o valor do limiar.

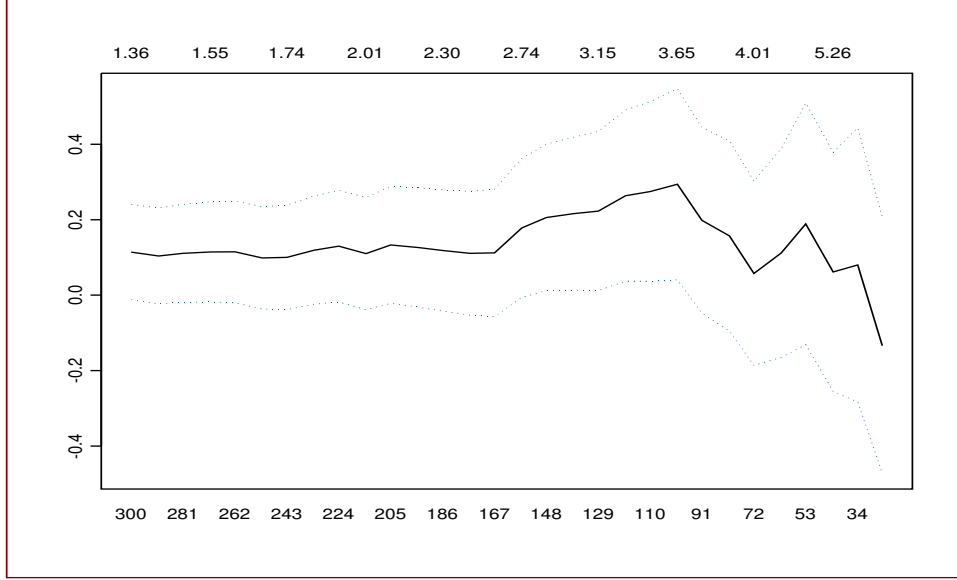


Figura 2.13: Sensibilidade das estimativas por máxima verossimilhança de ξ em relação à escolha do limiar u , para a cauda esquerda do índice Brasil. No eixo horizontal o número de excessos usados no processo de estimação.

Para um ajuste GPD sugerimos utilizar-se várias ferramentas para a escolha do limiar. Entre outras, sugerimos examinar as funções $e_N(u)$ e $e_N^*(u)$; fazer um histograma dos dados de excessos e examinar o decaimento da cauda, a qual deve ser estritamente decrescente; utilizar o gráfico da Figura 2.13.

Os erros padrões e intervalos de confiança para as estimativas dos parâmetros da GPD podem ser construídos através do método bootstrap, embora aqui a suposição de independência (dos excessos) seja mais difícil de ser defendida. Assim como no caso do ajuste GEV, deve-se verificar a qualidade do ajuste através de um teste Kolmogorov formal, e também graficamente. Gráficos utilizando a função de distribuição empírica dos dados, histogramas, densidade e f.d.a. ajustadas, e QQ-Plots, devem ser examinados. Pode-se testar a hipótese nula $H_0: \xi = 0$ fazendo o teste da razão de verossimilhanças, conforme explicado no caso GEV. Aqui o modelo completo é GPD e o restrito (H_0) é o Exponencial.

MGPD: Uma Extensão da GPD. Embora a GPD pareça uma boa candidata para modelar observações de cauda, é possível usar um modelo ainda mais poderoso, proposto por Anderson, C. W. e Dancy, G. P. (1992): a distribuição de Pareto Generalizada Modificada (MGPD).

A MGPD tem função de distribuição dada por

$$P_{\gamma, \xi, \psi}(y) = \begin{cases} 1 - (1 + \frac{\xi y^\gamma}{\psi})^{-1/\xi} & \text{se } \xi \neq 0 \\ 1 - \exp(-\frac{y^\gamma}{\psi}) & \text{se } \xi = 0 \end{cases} \quad (2.42)$$

onde $\gamma \in \mathbb{R}$. Notemos que para $\gamma = 1$ temos o caso restrito da família GPD. Para uma aplicação em finanças veja Mendes, B. V. M. e Brandi, V. R. (2004).

2.11 Apêndice do Capítulo 2

AP2.1: Convergência em probabilidade e convergência quase certa. Uma sequência de v.a.'s $\{X_n, n \geq 1\}$ em um espaço de probabilidade (Ω, F, P) converge em probabilidade para X , $X_n \xrightarrow{p} X$, se existe uma v.a. X em (Ω, F, P) tal que

$$Pr\{|X_n - X| > \epsilon\} \rightarrow 0, \quad \forall \epsilon > 0, \quad \text{quando } n \rightarrow \infty.$$

E, se existir uma v.a. X em (Ω, F, P) tal que

$$Pr\{\lim_{n \rightarrow \infty} X_n = X\} = 1$$

teremos convergência quase certa ($X_n \xrightarrow{q.c.} X$).

O lema 1 da página 66 de Chow, Y. S. e Teicher, H. (1988a) estabelece que

Lema 1. Variáveis aleatórias $X_n \xrightarrow{q.c.} X$, X uma v.a., se e somente se $\sup_{j \geq n} |X_j - X| \xrightarrow{p} 0$ se e somente se para todo $\epsilon, \delta > 0$

$$Pr\left\{\bigcap_{j=n}^{\infty} [|X_j - X| \leq \epsilon]\right\} \geq 1 - \delta \quad \text{para } n \geq N(\epsilon, \delta).$$

AP2.2: Distribuições do mesmo tipo. Duas distribuições F e F^* são do mesmo tipo se existirem constantes a e b tais que

$$F^*(ax + b) = F(x) \quad \forall x.$$

Exemplo: Suponha $Z \sim F = N(0, 1)$, normal padrão, e $Z^* = aZ + b$. Então $Z^* \sim F^* = N(b, a^2)$ e $F^*(ax + b) = F(x)$.

AP2.3: Domínio de atração. Veremos algumas condições para que uma distribuição F pertença ao domínio de atração de uma Gumbel, de uma Fréchet, de uma Weibull. As condições dadas a seguir são necessárias e suficientes. Foram obtidas por Gnedenko. B. V. (1943), onde foi também provado que as três distribuições são as únicas satisfazendo a condição de max-estabilidade.

Seja X_q o quantil q da distribuição de X , isto é, $Pr\{X \leq X_q\} = q$, ou $F(X_q) = q$. As condições são

$$\lim_{n \rightarrow \infty} n[1 - F(X_{1-\frac{1}{n}} + a(X_{1-\frac{1}{ne}} - X_{1-\frac{1}{n}}))] = e^{-a}$$

para o tipo I.

Para o tipo II é

$$\lim_{x \rightarrow \infty} \frac{1 - F(x)}{1 - F(cx)} = c^k, \quad c > 0, \quad k > 0$$

Para o tipo III é

$$\lim_{x \rightarrow o^-} \frac{1 - F(cx + x^+)}{1 - F(x + x^+)} = c^k, \quad c > 0, \quad k > 0$$

onde x^+ é o limite superior da distribuição de X .

AP2.4: Variação regular no infinito.

- (i) Uma função positiva L em $(0, \infty)$, mensurável a Lebesgue, é de variação lenta no infinito se $\lim_{x \rightarrow \infty} \frac{L(tx)}{L(x)} = 1, t > 0$.
- (ii) Uma função positiva h em $(0, \infty)$, mensurável a Lebesgue, é de variação regular no infinito de índice $\alpha \in \mathfrak{R}$ se $\lim_{x \rightarrow \infty} \frac{h(tx)}{h(x)} = t^\alpha, t > 0$.

Por exemplo, para $\alpha \in \mathfrak{R}$ as funções x^α , $x^\alpha(\ln(1+x))$, e $(x(\ln(1+x)))^\alpha$ são de variação regular no infinito de índice α .

AP2.5: Inversa Generalizada. A inversa generalizada F^\leftarrow de uma função F não decrescente em \mathfrak{R} é definida como

$$F^\leftarrow(t) = \inf\{x \in \mathfrak{R} : F(x) \geq t\}.$$

Quando F é uma f.d. a inversa generalizada de F é chamada de função quantil, e $x_p = F^\leftarrow(p)$ define o quantil p (ou p -quantil) de F .

A função quantil da GEV padrão ($\mu = 0, \sigma = 1$) é, para $\xi \neq 0$

$$H_\xi^{-1}(q) = ((-\ln(q))^{-\xi} - 1)/\xi.$$

AP2.6: Estimação pelo Método dos Momentos. Sejam X, X_1, X_2, \dots, X_n v.a.s aleatórias i.i.d. com f.d. F indexada por um parâmetro $\theta \in \mathfrak{R}^p$. O Método dos Momentos consiste em montar p equações tendo de um lado os momentos populacionais de F , e do outro os momentos amostrais. Os momentos populacionais são $E[X], E[X^2], \dots, E[X^p]$; e os momentos amostrais são $\frac{1}{n} \sum_{i=1}^n X_i^1, \frac{1}{n} \sum_{i=1}^n X_i^2, \dots, \frac{1}{n} \sum_{i=1}^n X_i^p$. As equações devem ser resolvidas para θ . Por exemplo, se F é a distribuição Normal com média μ e variância σ^2 , desconhecidos, temos que $p = 2, \theta = (\mu, \sigma^2)$, e basta montar as duas equações

$$E[X] = \mu = \frac{1}{n} \sum_{i=1}^n X_i^1$$

e

$$E[X^2] = \sigma^2 + \mu^2 = \frac{1}{n} \sum_{i=1}^n X_i^2,$$

para obter $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i$, e $\widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^n X_i^2 - \hat{\mu}^2$.

AP2.7: Momentos, Densidade e Função Quantil da $P_{\xi, \psi}$. Nos casos particulares Exponencial (I), Pareto (II) e Beta (III), padrão, temos

$$\eta_{j, P_{II, \alpha}} = E[Y^j] = \frac{\alpha}{\alpha - j} \quad \text{se } \alpha > j$$

$$\eta_{j,P_{III},\alpha} = E[Y^j] = (-1)^j \frac{\alpha}{\alpha - j}$$

Notemos que o j -ésimo momento de uma Pareto padrão é infinito se $j \geq \alpha$. Se $\alpha > 2$, as variâncias dos tipos *II* e *III* padrão são dadas por

$$\text{var}_{P_{i,\alpha}} = \frac{\alpha}{(\alpha - 1)^2(\alpha - 2)} \quad i = II, III .$$

A GPD padrão (locação zero e escala um), parametrizada por ξ , tem esperança e variância

$$\eta_{P_\xi} = E[Y^1] = \frac{1}{1 - \xi} \quad , \quad \xi < 1$$

$$\text{var}_{P_\xi} = \text{var}(Y) = \frac{1}{(1 - \xi)^2(1 - 2\xi)} \quad \text{se } \xi < 1/2$$

Se $X \sim P_{\xi,\psi}$, então para $\xi < \frac{1}{k}$ com $k \in \mathcal{N}$ temos

$$E[X^k] = \frac{\psi^k \Gamma(\frac{1}{\xi} - k) k!}{\xi^{k+1} \Gamma(1 + \frac{1}{\xi})}$$

Por exemplo, a esperança de X , $E[X]$, para $\xi < 1$ é

$$E[X] = \frac{\psi}{1 - \xi} .$$

A densidade da $P_{\xi,\psi}$, denotada por $p_{\xi,\psi}(y)$, é

$$\begin{cases} \frac{1}{\psi} (1 + \xi \frac{y}{\psi})^{-(1+1/\xi)} & \text{para } 0 \leq y, \quad \text{se } \xi \neq 0 \\ \frac{1}{\psi} \exp\{-\frac{y}{\psi}\} & \text{para } 0 \leq y \leq \frac{1}{|\xi|}, \quad \text{se } \xi = 0 . \end{cases} \quad (2.43)$$

A função quantil é, para $\xi \neq 0$

$$H_\xi^{-1}(q) = \frac{1}{\xi} ((1 - q)^{-\xi} - 1) .$$

AP2.8: Função Média Podada dos Excessos. É outra versão robusta da $e(u)$. Sua motivação surge do fato que a $e_{N_u}(u)$ é um estimador bastante impreciso de $e(u)$ se $\alpha > 1$, porém perto de 1. Lembremos que a f.d. Pareto não existe para $\alpha < 1$.

Sua definição lembra a de estimadores podados, como, por exemplo, a média podada. Para $0 < p \leq 1$, é definida como

$$e_{F_u,p}(u) = \frac{1}{p} \int_{-\infty}^{F_u^{-1}(p)} x dF_u(x) .$$

Onde temos que: $F_u^{-1}(p)$ é o p -quantil da distribuição condicional de $X - u | X > u$; Não ocorre “poda” (retirada de dados) se $p = 1$; O fator $\frac{1}{p}$ normaliza a distribuição; A $e_{F_u,p}(u)$ da GPD é também linear; Para o tipo *II* é estritamente crescente para cada α , ($\alpha > 0$); Para o tipo *III* é

estritamente decrescente; Para o tipo I é constante; A versão empírica é obtida substituindo a f.d. F_u pela sua versão empírica.

AP2.9: Teste da razão das verossimilhanças (Bickel, P.J. e Doksum, K.A. (2001)). Suponha que as v.a.'s $\mathbf{r} = (r_1, r_2, \dots, r_T)$ tenham densidade ou função de probabilidade $f(\mathbf{r}, \theta)$, que $f(\mathbf{r}, \theta)$ seja uma função contínua de θ , e que desejamos testar $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$, sendo Θ_0 de dimensão menor que $\Theta = \Theta_0 \cup \Theta_1$. A estatística teste do teste da razão das verossimilhanças é dada por

$$\lambda(\mathbf{r}) = \frac{\sup\{f(\mathbf{r}, \theta) : \theta \in \Theta\}}{\sup\{f(\mathbf{r}, \theta) : \theta \in \Theta_0\}}.$$

Para efetuar o teste devemos (a) Calcular o EMV $\hat{\theta}$ de θ ; (b) Calcular o EMV $\hat{\theta}_0$ de θ quando θ varia apenas em Θ_0 ; (c) Calcular $\lambda(\mathbf{r}) = \frac{f(\mathbf{r}, \hat{\theta})}{f(\mathbf{r}, \hat{\theta}_0)}$; (d) Encontrar uma função $h()$ estritamente crescente com λ e tal que $h(\lambda(\mathbf{r}))$ tenha uma forma simples e possa ser tabelada sob H_0 .

O caso mais geral ocorre quando as r_i 's são independentes, θ é um vetor paramétrico de dimensão l variando sobre um conjunto aberto Θ , o *mapping* $\theta \rightarrow f(\mathbf{r}, \theta)$ é suave, e o espaço paramétrico Θ_0 sob a hipótese nula pode ser escrito como $\Theta_0 = \{\theta \in \Theta : \theta = \theta(\eta), \eta \in E\}$, onde o *mapping* $\eta \rightarrow \theta(\eta)$ é suave e E é um conjunto aberto de um espaço de dimensão q . Seja $\lambda(\mathbf{r})$ a estatística do teste da razão das verossimilhanças para testar $H_0 = \Theta_0$ versus $H_1 = \Theta_1 = \Theta - \Theta_0$.

Sob algumas outras condições de regularidade, sob a hipótese nula e para T grande, $2 \log \lambda(\mathbf{r})$ tem distribuição aproximadamente quiquadrado com $l - q$ graus de liberdade, χ^2_{l-q} . O teste de nível α é unilateral para a direita e rejeita para valores da estatística teste maiores que o seu quantil de probabilidade acumulada $(1 - \alpha)$.

2.12 Exercícios

Exercício 1: Expressar a esperança e a variância de r_t em função dos momentos $E[R_t]$ e $\text{var}(R_t)$, veja as formulas (2.5) e (2.6).

Exercício 2: Faça a figura sugerida em (2.1.7).

Exercício 3: Seja (X_i) uma seqüência de v.a.s i.i.d. exponenciais-padrão, e M_n o $\max(X_1, \dots, X_n)$. Calcule o $\lim_{n \rightarrow \infty} Pr\{\frac{M_n - b_n}{a_n} \leq x\}$ para $a_n = 1$ e $b_n = \ln(n)$.

Exercício 4: Seja (X_i) uma seqüência de v.a.s i.i.d. Cauchy padrão, isto é, $f_X = \frac{1}{\pi(1+x^2)}$, $x \in \mathbb{R}$, e M_n o $\max(X_1, \dots, X_n)$. Seja $a_n = 1$ e $b_n = \pi/n$. Mostre que a distribuição-limite do máximo normalizado é Fréchet com $\alpha = 1$. Dica: Usando a regra de L'Hospital obtemos, $\lim_{n \rightarrow \infty} \frac{\bar{F}_X(x)}{(\pi x)^{-1}} = 1$.

Exercício 5: Gere 1.000 v.a.s $X \sim \text{Exponencial}(1)$. Colete M_n para $n = 3$. Normalize M_n usando $a_n = 1$ e $b_n = \ln(n)$. Faça a f.d. empírica dos 1.000 dados, e no mesmo gráfico coloque a f.d. de uma Gumbel. Comente.

Exercício 6: Repetir o exercício 5 para $n=5$ e $n=10$. Compare os resultados.

Exercício 7: Repita os exercícios 5 e 6 supondo $X \sim N(0, 1)$. Compare com os resultados obtidos nos exercícios 4 e 5.

Exercício 8: Mostre que o limite superior de uma Weibull é dado por

$$x^+ = \mu + \frac{\sigma}{\xi}.$$

Exercício 9: Um resultado de Gnedenko estabelece que $F \in MDA(G_\xi)$, $\xi > 0 \iff 1 - F(x) = x^{-1/\xi} L(x)$, onde L é tal que $\frac{L(tx)}{L(t)} \rightarrow 1$, $t \rightarrow \infty$, $\forall x$. Vale então que

$$\ln(1 - F(x)) \sim \text{constante} - \frac{1}{\xi} \ln(x).$$

Este resultado serve como base para uma análise exploratória. Usando F_T (f.d.a.e.) no lugar de F em $1 - F(x) = x^{-1/\xi} L(x)$, faça o gráfico de $\ln(1 - F_T(x_{(i)}))$ vs $\ln(x_{(i)})$. Uma reta com inclinação negativa indicaria caudas pesadas, $\xi > 0$. Usar dados de cauda, além de um limiar.

Exercício 10: Procure dados de temperaturas ou níveis de chuva etc, de preferência anuais. Procedimentos exploratórios: Visualizar os dados através de estimativas da densidade: (i) ajuste não paramétrico; (ii) ajuste de uma densidade Gumbel. Visualizar os dados através de plot tipo série temporal (ou dispersão). Existe tendência? Ajuste uma reta.

Exercício 11: Obtenha 10 séries de retornos financeiros, incluindo ações brasileiras e americanas, índices de países desenvolvidos e emergentes etc. Calcule a estimativa de $\xi = \frac{1}{\alpha}$ usando (2.41) para essas séries fixando r de tal maneira que use os maiores 10% dos dados para estimação. Compare seus resultados com a estimação via regressão robusta.

Exercício 12: Para as mesmas séries do exercício anterior varie o valor de r . Escolha três delas e faça o gráfico de r versus ξ .

Exercício 13: Para uma das séries de retornos financeiros, e para as perdas e para os ganhos, verifique a estabilidade das estimativas de L-momentos dos parâmetros da GEV, em particular de ξ , em relação ao tamanho n do bloco. Isto é, obtenha os extremos semanais, quinzenais, mensais, bimestrais, semi-anuais e anuais; estime ξ , μ , e σ para esses conjuntos de dados. Faça gráficos ou tabelas com os resultados.

Exercício 14: Repita o exercício anterior agora para uma série de 2.000 observações geradas de uma Fréchet com $\mu = 0$, $\sigma = 1$ e $\xi = 0.4$. Obtenha as estimativas de L-momentos, varie o tamanho do bloco n , e observe a qualidade de suas estimativas em relação ao tamanho do bloco.

Exercício 15: Construa intervalos de confiança pelo método bootstrap para os parâmetros da $G_{\xi,\mu,\sigma}$ dos mínimos mensais de uma das séries de retornos financeiros do Exercício 10.

Exercício 16: Faça os testes estatísticos e análises gráficas estudadas para verificar a qualidade do ajuste para uma das séries de retornos financeiros do Exercício 10. Faça a comparação da f.d. empírica e f.d. ajustada, superpostas em um mesmo gráfico, e a comparação da densidade não paramétrica e da densidade estimada, superpostas ao histograma.

Exercício 17: Gere quatro séries de tamanho $N = 500$ da distribuição GPD com parâmetros $\xi = -0, 4, -0, 1, 0, 0, 0, 1, 0, 4$, respectivamente. Fixe a escala em $\psi = 1$. Para as quatro séries faça o gráfico de $e_N(u)$, de $e_N^*(u)$ e da Média dos Excessos Podada (dada no Apêndice [AP2.8]). Verifique se as características teóricas dessas funções se verificam.

Exercício 18: Contamine cada série do exercício anterior com 5 (1%) valores bem extremos e novamente faça o gráfico das três funções. Comente sobre a robustez (ou falta de) dessas funções.

Capítulo 3

Modelagem Não Condicional Multivariada

Modelos financeiros são essencialmente multivariados. Como um exemplo simples temos que as variações do preço de uma carteira depende de um conjunto de d ativos. Neste capítulo veremos como modelar o vetor d -dimensional \mathbf{r} de uma forma estática, sem considerar suas características temporais. Isto é, iremos estimar a distribuição conjunta $F_{\mathbf{r}}$ de \mathbf{r} , a qual especifica não só as distribuições marginais F_i , $i = 1, \dots, d$, mas também a estrutura de dependência ligando as marginais.

Considere uma amostra de tamanho T dos log-retornos correspondentes a d ativos:

$$\{r_{it}; i = 1, \dots, d; t = 1, \dots, T\}.$$

Sua distribuição conjunta é

$$F_{\mathbf{r}}((r_{11}, \dots, r_{d1}), (r_{12}, \dots, r_{d2}), \dots, (r_{1T}, \dots, r_{dT}); \mathbf{Y}, \theta),$$

onde $\mathbf{r} \in \mathbb{R}^d$, \mathbf{Y} é um vetor de variáveis (por exemplo, econômicas) relacionadas com \mathbf{r} , e θ é o vetor de parâmetros desconhecidos. Esta distribuição $F_{\mathbf{r}}$ governa o comportamento estocástico dos retornos e de \mathbf{Y} .

Por sua importância como distribuição central limite para a qual converge a distribuição de muitos estimadores, e também por sua facilidade matemática, a distribuição Normal multivariada quase sempre é o único modelo conhecido assumido por analistas quantitativos para representar $F_{\mathbf{r}}$. Uma maior flexibilidade pode ser obtida assumindo-se uma distribuição elíptica multivariada, a qual generaliza a distribuição Normal retendo várias de suas importantes propriedades. No caso de uma distribuição elíptica a estrutura de dependência, ou estrutura de correlação, é obtida a partir da matriz Σ . As distribuições esféricas e elípticas e seus casos particulares importantes, Normal e t_{ν} , foram definidas no Capítulo 2.

Embora assumida, raramente a distribuição Normal multivariada é confirmada pelos testes estatísticos formais e/ou inspeções visuais. Se $\mathbf{r} \sim N_d(\boldsymbol{\mu}, \Sigma)$, então cada distribuição marginal F_i ,

$i = 1, \dots, d$ deve ser Normal. Assim, para testar normalidade multivariada, podemos começar testando se cada uma das marginais é Normal, com os testes dados nos capítulos anteriores ou através de um qq-plot. Contudo, conforme veremos na Seção 3.1, podemos ter margens Normais ligadas por outra estrutura de dependência que não a da distribuição Normal, descaracterizando a Normal multivariada.

Para testar normalidade conjunta considere a amostra $T \times d$ de retornos e a estatística teste

$$\{D_i^2 = (\mathbf{r}_i - \bar{\mathbf{r}})' \mathbf{S}^{-1} (\mathbf{r}_i - \bar{\mathbf{r}})\}$$

para $i = 1, \dots, T$, onde \mathbf{r}_i representa a i -ésima linha da matriz de dados. Os D_i não são independentes pois são definidos a partir do vetor de médias amostrais e do estimador para a covariância. A distribuição de D_i^2 sob a hipótese nula não é exatamente uma chi-quadrado, mas pode ser aproximada por ela para T grande. Temos então que $D_1^2, D_2^2, \dots, D_T^2$ é considerada uma amostra de uma v.a. χ_d^2 e esta suposição pode ser verificada através de um qq-plot ou um teste KS.

Conhecer a distribuição multivariada é importante quando queremos, por exemplo, calcular a variabilidade de alguma medida (de risco) através de simulações. Embora bastante promissoras, as distribuições elípticas assimétricas apresentam a mesma grande restrição de requerer que suas marginais univariadas sejam do mesmo tipo, o que não ocorre na prática. Em geral dados multivariados se caracterizam por terem suas componentes univariadas possuindo as mais diversas distribuições, por exemplo, uma delas pode ser discreta, outra tipo Bernoulli, outra simétrica, outras assimétricas, etc, impossibilitando sua modelagem adequada com qualquer um dos modelos multivariados elípticos.

Além da correta especificação das distribuições univariadas, a estimação e a simulação de muitos produtos em finanças requer uma modelagem acurada da dependência entre os ativos. Para capturar todos os tipos de dependência, linear e não-linear, é preciso usar cópulas. Esta afirmação justifica-se por várias razões, principalmente pela inexistência de uma classe de distribuições multivariadas flexível bastante para incluir assimetrias, modelar dependência linear e não-linear, e admitir que suas margens possuam distribuições completamente diferentes. A modelagem através de cópulas permite que se fatore a distribuição multivariada $F_{\mathbf{r}}$ nas suas marginais univariadas e na função cópula. Na prática, este fato simplifica e sofisticas as fases de *especificação*, *estimação* e *simulação* da *distribuição multivariada*, veja expressão (3.6).

3.1 Cópulas

A partir do acordo internacional da Basiléia II observou-se uma certa motivação para o desenvolvimento de ferramentas estatísticas mais sofisticadas em finanças. Por detrás de cada ferramenta temos a suposição de um modelo probabilístico e, por muito tempo, a modelagem em finanças somente considerou a distribuição Normal multivariada. Isto deve-se parcialmente ao fato de que a maioria dos resultados importantes nesta área baseavam-se na suposição de normalidade, mas também deve-se à falta de alternativas para a distribuição multivariada e às limitações computacionais. Contudo, uma simples análise exploratória feita em qualquer coleção de log-retornos de índices, ações, carteiras ou títulos revelam desvios significativos da normalidade.

Um conjunto de log-retornos em diversos ativos em geral possuem duas características especiais: (I) cada margem tipicamente tem seu próprio grau de assimetria e alta curtose, assim como alguns padrões específicos de dinâmica temporal, e (II) a estrutura de dependência entre os diversos pares varia substancialmente, indo desde independência até formas complexas de dependência não linear. *Nenhuma família natural de distribuição multivariada, por exemplo, a família elíptica, inclui essas características.* Isto é verdade na modelagem não condicional (tópico deste capítulo) mas também se aplica à distribuição dos erros de modelos dinâmicos sofisticados.

A solução é o uso de cópulas, introduzidas por Sklar, A. (1959), estudadas nos livros de Joe, H. (1997) e Nelsen, R.B. (2006), e aplicadas na área de finanças por Embrechts, P., McNeil, A. J. e Straumann, D. (2002). Desde então cópulas tem sido usadas para modelar um conjunto de ativos (Mendes, B. V. M. e Souza, R. M. (2004) e Mendes, B. V. M., Semeraro, M. M. e Leal, R. P. C. (2010)), *hedge funds*, probabilidades conjuntas de *default*, etc.

As áreas da atuária e de finanças são as que mais se beneficiaram dos avanços em cópulas. Isto se deve muito ao fato de que a estimação (no tempo) das distribuições marginais univariadas pode ser feita com sucesso através de modelos da família GARCH e modelos da teoria dos valores extremos. Assim, como veremos mais a frente, resta apenas ajustar cópulas aos dados padronizados para se ter um ajuste de um modelo multivariado. Para isto um número incontável de famílias de cópulas paramétricas estão ao nosso dispor. Mas é preciso conhecer as características de cada família, o tipo de dependência que ela modela ou prioriza, para escolher bem aquela a ser ajustada aos dados.

Cópula é uma distribuição multivariada com distribuições marginais uniformes padrão.

Existem outras maneiras de se definir uma cópula. Para um inteiro positivo d , denotemos o d -cubo unitário por I^d . Uma cópula é qualquer função $C : I^d \rightarrow [0, 1]$ possuindo as seguintes propriedades:

1. C é *grounded*, isto é, para todo x_i em $[0, 1]$ e todo k em $[1, d]$:

$$C(x_1, \dots, x_{k-1}, 0, x_{k+1}, \dots, x_d) = 0 \quad (3.1)$$

2. C possui marginais uniformes, isto é, para todo x em $[0, 1]$:

$$C(1, \dots, 1, x, 1, \dots, 1) = x \quad (3.2)$$

3. C é d -crescente, isto é, seja $I_k = (a_k, b_k]$, $a_k < b_k$, $k = 1, 2, \dots, d$. Seja $g : \mathbb{R}^k \rightarrow \mathbb{R}$, e definimos: $\Delta_{I_k}^k g(x_1, x_2, \dots, x_k) = g(x_1, x_2, \dots, x_{k-1}, b_k) - g(x_1, x_2, \dots, x_{k-1}, a_k)$. Então, para todo I_k e para $k = 1, \dots, d$ tem-se:

$$\Delta_{I_1}^1 \cdots \Delta_{I_d}^d C(x_1, \dots, x_d) \geq 0. \quad (3.3)$$

As condições (3.1) e (3.2) são geralmente chamadas de *condições de contorno*, e a (3.3) é a *condição de monotonicidade*. A condição de monotonicidade cresce em complexidade com o crescimento exponencial dos cantos do d -cubo, e é obviamente a condição mais difícil de se verificar.

Podemos também definir uma cópula d -variada, e agora em particular a bivariada, como uma função $C : [0, 1]^2 \mapsto [0, 1]$, responsável por ligar as distribuições marginais univariadas F_1 e F_2 para

formar sua distribuição bivariada F . Assim, C contém toda a informação sobre a estrutura de dependência entre F_1 e F_2 . Por simplicidade, a partir deste ponto iremos dar os conceitos no plano bivariado.

Uma ferramenta importante em aplicações é o teorema de Sklar. De acordo com o teorema, para toda função de distribuição (f.d.a.) bidimensional F com marginais contínuas F_1, F_2 , existe uma única cópula. A importância deste resultado é que ele permite que se estude a estrutura de dependência de uma distribuição multivariada sem estudar as distribuições marginais.

Teorema de Sklar: Seja F uma f.d.a. conjunta em \mathbb{R}^2 com marginais $F_i, i = 1, 2$. Então, existe uma cópula C tal que $\forall \mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$:

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2)) . \quad (3.4)$$

Quando todas F_i são contínuas, a C é única. Reciprocamente, se C é uma cópula e as F_i s são f.d.a., então a função F definida por (3.4) é uma f.d.a. conjunta com marginais F_i . Quando F não é contínua, ainda existe a representação através de uma cópula, mas ela não é mais única.

A equação (3.4) pode ser invertida da seguinte maneira. Seja F a f.d.a. conjunta do vetor aleatório contínuo $\mathbf{X} = (X_1, X_2)$. Sejam F_1, F_2 as f.d.a. marginais, e definimos $X_i \mapsto F_i(X_i) = U_i, i = 1, 2$, como a transformação integral da probabilidade para a distribuição Uniforme(0,1). Então, a cópula C pertinente a F é definida por

$$C(u_1, u_2) = F(F_1^{-1}(u_1), F_2^{-1}(u_2)), \quad (3.5)$$

$\forall (u_1, u_2) \in (0, 1)^2$, e onde F_i^{-1} , é a função quantil de $F_i, i = 1, 2$, isto é, $F_i^{-1}(p) = \inf\{x | F_i(x) \geq p\}, p \in (0, 1)$.

Um outro teorema de particular importância é:

(Teorema 2.2.7 de Nelsen [1998]). Seja C uma cópula. Para todo u_2 em $[0, 1]$, a derivada parcial $\frac{\partial C}{\partial u_1}$ existe para quase todo u_1 , e para tais u_2 e u_1 ,

$$0 \leq \frac{\partial C(u_1, u_2)}{\partial u_1} \leq 1 .$$

Similarmente, para todo u_1 em $[0, 1]$, a derivada parcial $\frac{\partial C}{\partial u_2}$ existe para quase todo u_2 , e para tais u_1 e u_2 ,

$$0 \leq \frac{\partial C(u_1, u_2)}{\partial u_2} \leq 1 .$$

Mais ainda, as funções $u_1 \mapsto \frac{\partial C(u_1, u_2)}{\partial u_2}$ e $u_2 \mapsto \frac{\partial C(u_1, u_2)}{\partial u_1}$ são definidas e não-decrescentes em quase toda a parte em $[0, 1]$.

A partir de (3.4) e do teorema acima obtemos que a densidade multivariada de \mathbf{X} é dada por

$$f(x_1, x_2) = c(F_1(x_1), F_2(x_2)) \cdot \prod_{i=1}^2 f_i(x_i) , \quad (3.6)$$

onde $c(u_1, u_2) = \partial C^2(u_1, u_2) / \partial u_1 \partial u_2$.

Na representação (3.6) fica clara a decomposição da densidade conjunta $f(x_1, x_2)$ em duas partes: uma que descreve apenas a estrutura de dependência e outra que descreve o comportamento marginal de cada variável. Percebemos então que a cópula de um vetor de variáveis aleatórias pode ser interpretada como a *estrutura de dependência* do vetor aleatório.

Existem cópias de grande interesse teórico. Elas são a cópula produto (de v.a.s independentes), e as cópias correspondentes aos limites (inferior e superior) de Fréchet-Hoeffding, as cópias W e M .

(i) *Cópula de marginais independentes* $C_\perp : C_\perp(u_1, u_2) = u_1 \cdot u_2$.

(ii) *Cópias de Fréchet*:

$$W(u_1, u_2) = \max\{(u_1 + u_2 - 1), 0\} ,$$

$$M(u_1, u_2) = \min\{u_1, u_2\} .$$

Elas estão relacionadas com a importante relação de ordem, que se verifica para toda cópula C e todo $(u_1, u_2) \in [0, 1]^2$

$$W(u_1, u_2) \leq C(u_1, u_2) \leq M(u_1, u_2) ,$$

e denotada por $W \prec C \prec M$. A interpretação das cópias limite pode ser feita da seguinte maneira: Se duas variáveis aleatórias possuem cópula M (W), então cada uma é quase certamente uma função crescente (decrescente) da outra.

A cópula de qualquer F captura e resume a força e tipo da dependência entre as variáveis, mesmo quando elas tenham sido objeto de transformações monótonas crescentes. Este resultado pode ser muito útil em algumas situações nas quais seja mais conveniente trabalhar com transformações crescentes das variáveis observadas e quando se deseja investigar a estrutura de dependência de (X_1, \dots, X_d) . Esta propriedade de invariância garante, por exemplo, que retornos financeiros e seus logaritmos tenham a mesma cópula. Resumindo, se (X_1, \dots, X_d) possui cópula C e T_1, \dots, T_d são funções contínuas crescentes, então $(T_1(X_1), \dots, T_d(X_d))$ também tem cópula C .

Algumas medidas de dependência possuem esta propriedade de invariância e ficam convenientemente representadas através da cópula C . Exemplos de tais medidas “baseadas em cópias”, são o coeficiente de Kendall (τ) e o de Spearman (ρ_S) definidos no Capítulo 1. São dados por:

$$\tau(X_1, X_2) = 4 \int \int_{[0,1]^2} C(u, v) dC(u, v) - 1 ,$$

e

$$\rho_S(X_1, X_2) = 12 \int \int_{[0,1]^2} C(u, v) dudv - 3 .$$

O coeficiente de correlação ρ_S baseado nos postos sempre existe no intervalo $[-1, 1]$, e *não* depende das distribuições marginais. Os valores ± 1 ocorrem quando as variáveis são funcionalmente dependentes, isto é, quando elas são modeladas por uma das cópias limite de Fréchet. Note que

$$\rho_S(X_1, X_2) = \rho(F_1(X_1), F_2(X_2)),$$

de tal maneira que no ambiente das cópias o coeficiente de Spearman e o linear coincidem.

Até recentemente, o coeficiente correlação linear de Pearson, ρ , era a única medida usada para medir a associação entre produtos financeiros. Apesar de ser a medida canônica de dependência no

mundo Gaussiano, ρ não possui esta propriedade de invariância (ser *copula based*), e depende também das distribuições marginais. Não podemos esquecer que ρ possui a deficiência de medir apenas a correlação *linear*, apresentando ainda outros problemas examinados em Embrechts, P., McNeil, A. J. e Straumann, D. (2002).

Outro importante conceito de dependência baseado na cópula é o coeficiente de dependência de cauda, λ_U , definido no Capítulo 1. Seja C a cópula de (X_1, X_2) . Segue que

$$\lambda_U = \lim_{u \uparrow 1} \frac{\overline{C}(u, u)}{1 - u}, \text{ onde } \overline{C}(u_1, u_2) = \Pr\{U_1 > u_1, U_2 > u_2\} \text{ e } \lambda_L = \lim_{u \downarrow 0} \frac{C(u, u)}{u},$$

se os limites existirem.

Dizemos que existe dependência assintótica na cauda superior (inferior) se $\lambda_U \in (0, 1]$ ($\lambda_L \in (0, 1]$). O λ_U (λ_L) mede a quantidade de dependência no extremo quadrante superior (inferior) da distribuição bivariada. Em finanças está relacionado com a força da associação durante eventos extremos. A cópula derivada ou pertinente à distribuição Normal multivariada não possui dependência de cauda. Assim, se esta cópula é assumida para modelar os log-retornos, ela irá subestimar os riscos conjuntos. Nem todas as cópulas possuem expressões fechadas para esses coeficientes, e algumas aproximações foram propostas na literatura.

Existem outros conceitos representando dependência de cauda, inclusive multivariados, veja Joe, H. (1996). Interessante observar que toda medida de dependência baseada na cópula é função crescente dos parâmetros da cópula. Na Tabela 3.1 damos 6 exemplos de famílias de cópulas que cobrem todos os tipos de dependência de cauda que os dados podem apresentar. Detalhes dessas e outras famílias, as expressões de suas densidades, e características podem ser vistos em Mendes, B. V. M. (2004).

Tabela 3.1: *Resumo das características de 6 cópulas selecionadas. Para a cópula t temos que $\lambda_L = \lambda_U = 2t_{\nu+1}(-\sqrt{\nu+1}\sqrt{1-\rho}/\sqrt{1+\rho})$. Para a cópula Gumbel, $\lambda_U = 2 - 2^{1/\theta}$. A cópula Clayton possui $\lambda_L = 2^{-1/\theta}$. Para a cópula BB7, $\lambda_U = 2 - 2^{1/\theta}$ o qual pode ser diferente de $\lambda_L = 2^{-1/\delta}$.*

Família	Tipo	Parâmetros	λ_L	λ_U	W	C_{\perp}	M
Gaussiana	Elíptica	$-1 \leq \rho \leq +1$	0.0	0.0	$\rho = -1$	$\rho = 0$	$\rho = +1$
t-copula	Elíptica	$-1 \leq \rho \leq +1, \nu$	$\sqrt{\nu}$	$\sqrt{\nu}$	$\rho = -1$	-	$\rho = +1$
Frank	Archim.	$-\infty < \theta < +\infty$	0.0	0.0	$\theta \rightarrow -\infty$	$\theta = 0$	$\theta \rightarrow +\infty$
Surv.Clayton	Archim./EV	$\theta \geq 0$	0.0	$\sqrt{\theta}$	-	$\theta = 1$	$\theta \rightarrow +\infty$
Gumbel	Archim./EV	$\theta \geq 1$	0.0	$\sqrt{\theta}$	-	$\theta = 1$	$\theta \rightarrow +\infty$
BB7	Archimedean	$\theta \geq 1, \delta > 0$	$\sqrt{\theta}$	$\sqrt{\delta}$	-	-	θ or $\delta \rightarrow \infty$

O conceito de dependência de cauda heterogênea, definida por Zhang, M. H. (2007), amplia o escopo de utilização da dependência de cauda na análise de dados. A dependência de cauda homogênea ocorre sobre a diagonal principal do quadrado $[0, 1]^2$, enquanto que a dependência de cauda heterogênea é calculada sobre a diagonal secundária e é determinada pelos cálculos dos coeficientes λ_{UL} e λ_{LU} , definidos abaixo. Considerando X_1, X_2, F_1 e F_2 , temos para a diagonal secundária as seguintes medidas:

Dependência de cauda heterogênea: λ_{UL} e λ_{LU} . O coeficiente da dependência de cauda superior-inferior é definido por

$$\lambda_{\text{UL}} := \lim_{q \rightarrow 1} P(X_2 < F_2^{(-1)}(1-q) \mid X_1 > F_1^{(-1)}(q)) \quad (3.7)$$

se este limite existe, e o coeficiente da dependência de cauda inferior-superior é definido por

$$\lambda_{\text{LU}} := \lim_{q \rightarrow 1} P(X_2 > F_2^{(-1)}(q) \mid X_1 < F_1^{(-1)}(1-q)). \quad (3.8)$$

As propriedades de dependência são importantes porque nos indicam se uma determinada cópula é adequada para a aplicação desejada. Uma apresentação exaustiva dos diversos tipos de associação entre v.a.s, incluindo conceitos de dependência positiva e negativa, conceitos de ordenação, medidas de dependência, medidas de concordância, etc., podem ser encontradas em Joe, H. (1997) ou em Nelsen, R.B. (2006). Também nessas mesmas referências veja as expressões das famílias paramétricas, com 1, 2, ou mais parâmetros.

A estimação de uma distribuição (paramétrica) multivariada pode ser afetuada em dois passos conforme sugerido pela equação (3.6). No primeiro passo são estimadas as distribuições marginais univariadas, obtendo as estimativas \hat{F}_{it} , $i = 1, 2$, $t = 1, \dots, T$, que são então usadas para calcular os valores pseudo-Uniforme(0,1), que serão utilizados no segundo passo para estimar os parâmetros da cópula.

A estimação marginal no primeiro passo pode ser feita baseada em algum dos modelos paramétricos dados no Capítulo 2, ou em um modelo semi-paramétrico, ou simplesmente utilizando a distribuição empírica. Alguns autores até preferem usar apenas a distribuição empírica para evitar o problema de especificação errônea (Frahm, G., Junker, M. e Schmidt, R., 2004). A cópula empírica é definida como: Seja $(X_{1,t}, X_{2,t})$, $t = 1, \dots, N$, T cópias independentes de (X_1, X_2) com distribuição conjunta F , marginais contínuas F_1 e F_2 , e cópula C . A *função de distribuição empírica* é dada por

$$F_T(x, y) = \frac{1}{T} \sum_{t=1}^T \mathbf{I}_{\{X_{1,t} \leq x, X_{2,t} \leq y\}}, \quad -\infty < x, y < +\infty,$$

onde $\mathbf{I}_{\{A\}}$ é a função indicadora. A *função de distribuição empírica marginal* $F_{i,T}(x)$, $i = 1, 2$, é definida como $F_{1,T} = F_T(x, +\infty)$ e $F_{2,T}(y) = F_T(+\infty, y)$. Seja $F_{i,T}^-$ representando a inversa generalizada de $F_{i,T}$, isto é, $F_{i,T}^-(u) = \sup\{t \in \mathbb{R} : F_{i,T}(t) \leq u, 0 \leq u \leq 1\}$. A *cópula empírica* \tilde{C} é definida por Fermanian, J. D., Radulovic, D. e Wegkamp, M. (2004)

$$\tilde{C}(u, v) = F_T(F_{1,T}^-(u), F_{2,T}^-(v)), \quad 0 \leq u, v \leq 1,$$

a qual estima sua distribuição não condicional.

No contexto de observações i.i.d., as estimativas dos parâmetros da cópula podem ser obtidas através do método da máxima verossimilhança (Joe, H., 1997), ou através de algum procedimento robusto combinado com estimadores de mínima distância (veja Tsukahara (2005) e Mendes, B. V. M., Melo, E. F. L. e Nelsen, R. B. (2007)), ou semi-parametricamente (Vandenhende, F. e Lambert, P., 2005). Neste último, uma nova família semi-paramétrica de cópulas é proposta e estimada usando-se o método de mínimos quadrados penalizados. Mendes, B. V. M. e Melo, E. F. (2009) e Mendes, B.

V. M. e Melo, E. F. (2010) propuzeram a estimação dinâmica para cópulas baseada em estimadores de máxima verossimilhança locais.

A boa qualidade do ajuste pode ser acessada visualmente através de pp-plots ou através de algum teste formal tipo GOF, usualmente baseados na minimização de algum critério e utilizando a cópula empírica. Veja exemplos de testes tipo GOF em Wang, W. e Wells, M. T. (2000), Breymann, W., Dias, A. e Embrechts, P. (2003), Chen, X., Fan, Y. e Patton, A. J. (2004), Genest, C., Quessy, J. F. e Remillard, B. (2006), e ainda o algoritmo PIT de Rosenblatt, M. (1952). A abordagem mais comumente empregada é a de transformar os dados em um conjunto de variáveis independentes Uniforme padrão e então calcular alguma medida de distância, como por exemplo a de Anderson-Darling ou a de Kolmogorov-Smirnov, entre as variáveis transformadas e a distribuição Uniforme. Para uma discussão de testes tipo GOF veja Remillard, B., Genest, C. e Beaudoin, D. (2009).

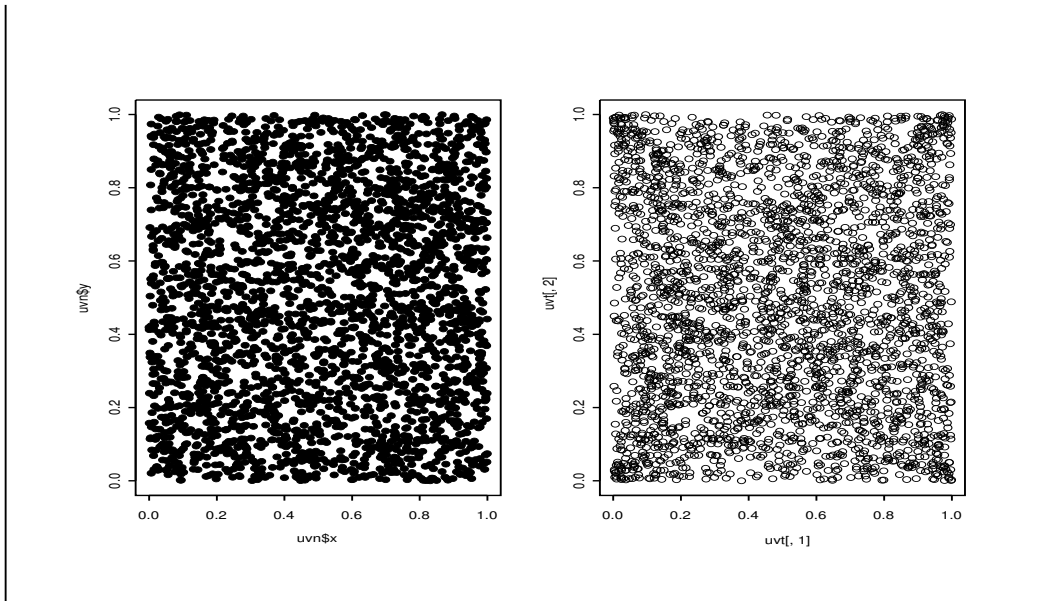


Figura 3.1: Valores simulados de uma cópula Gaussiana e uma *t*-student (4gl) com $\rho = 0.0$.

A Figura 3.1 mostra dados simulados das cópulas Gaussiana e *t*-student (4gl), ambas com $\rho = 0.0$. Observe que a cópula *t* possui dependência de cauda, mesmo para $\rho = 0.0$.

3.2 Pair-copulas

O uso de cópulas simplifica a especificação e estimação de uma distribuição multivariada. Conforme vimos, inicialmente as distribuições univariadas podem ser ajustadas a partir de uma grande quantidade de modelos disponíveis. No segundo passo, a estrutura de dependência entre as variáveis é modelada com uma cópula. Contudo, esta abordagem também tem suas limitações. Embora sejamos capazes de conseguir excelentes ajustes univariados (condicionais e não-condicionais), cada um es-

pecialmente adequado para cada marginal, na hora do ajuste da cópula existem alguns obstáculos relacionados com o problema de otimização envolvido, devido ao que se costuma chamar de “curse of dimensionality” (Scott, D. W., 1992). Infelizmente, a maioria dos pacotes computacionais disponíveis consegue lidar apenas com o caso bivariado. Mesmo quando somos capazes de ajustar uma cópula d -dimensional, $d > 2$, deparamos com o problema de que em geral as famílias de cópulas restringem que todos os pares tenham o mesmo tipo ou força de dependência. Por exemplo, no caso da cópula t , além do coeficiente de correlação, um único parâmetro, o número de graus de liberdade é utilizado para calcular o coeficiente de dependência de cauda para *todos* os pares, assim violando (II) (que a dependência entre os pares de variáveis varia substancialmente).

Pair-copulas, sendo uma coleção de potencialmente diferentes cópulas bivariadas, é uma abordagem flexível e interessante. O método de construção é hierárquico, onde as variáveis são sequencialmente incorporadas condicionalmente enquanto se move do nível 1 (*tree* 1) até o passo $d - 1$. As cópulas bivariadas componentes variam livremente, quanto à família e quanto aos valores dos parâmetros. Assim, todos os tipos e graus de dependência podem ser cobertos. Pair-copulas podem ser facilmente estimadas e simuladas, tornando-as uma abordagem muito apropriada para a modelagem em finanças. Como exemplos de aplicações em finanças veja Czado, C. e Min, A. (2009), Aas, K., Czado, C., Frigessi, A. e Bakken, H. (2009), Aas, K. e Berg, D. (2009), Fischer, M. J., Köck, C., Schlüter, S. e Weigert, F. (2007). Mendes, B. V. M. e Marquez, D. (2012) vai além da inferência e utiliza pair-copulas para a construção de fronteiras eficientes e para o cálculo do risco.

Considere um processo d -variado estacionário $(X_{1,t}, X_{2,t}, \dots, X_{d,t})_{t \in \mathcal{Z}}$, \mathcal{Z} um conjunto de índices. Quando a lei de distribuição conjunta de $(X_{1,t}, X_{2,t}, \dots, X_{d,t})$ é independente de t , a estrutura de dependência de $\mathbf{X} = (X_1, X_2, \dots, X_d)$ é dada por sua cópula (constante) C . Se \mathbf{X} é um vetor aleatório contínuo com f.d.a. F , densidade f , e f.d.a. marginais F_i com densidades f_i , $i = 1, 2, \dots, d$, então existe uma única cópula C pertinente a F , definida em $[0, 1]^d$ tal que

$$C(F_1(x_1), F_2(x_2), \dots, F_d(x_d)) = F(x_1, x_2, \dots, x_d) \quad (3.9)$$

se verifica para todo $(x_1, x_2, \dots, x_d) \in \mathfrak{R}^d$ (Sklar, 1959). Seja $F_i(X_i) \equiv U_i$, $i = 1, \dots, d$. Das suposições feitas, U_i segue uma Uniforme(0, 1). Tomando derivadas parciais obtemos

$$f(x_1, \dots, x_d) = c_{1\dots d}(F_1(x_1), \dots, F_d(x_d)) \prod_{i=1}^d f_i(x_i) \quad (3.10)$$

para alguma cópula d -dimensional com densidade $c_{1\dots d}$.

A decomposição de uma distribuição multivariada em uma cascada de cópulas foi originalmente proposta por Joe, H. (1997), e discutida em detalhes mais tarde por Bedford, T. e Cooke, R. (2001), Bedford, T. J. e Cooke, R. (2002), Kurowicka, D. e Cooke, R. M. (2006), e Aas, K., Czado, C., Frigessi, A. e Bakken, H. (2009).

Inicialmente notemos que toda densidade multivariada pode ser unicamente decomposta como

$$f(x_1, \dots, x_d) = f_d(x_d) \cdot f(x_{d-1}|x_d) \cdot f(x_{d-2}|x_{d-1}, x_d) \cdots f(x_1|x_2, \dots, x_d). \quad (3.11)$$

As densidades condicionais em (3.11) podem ser escritas como funções da densidade da cópula correspondente. Isto é, para todo j

$$f(x | v_1, v_2, \dots, v_d) = c_{xv_j|\mathbf{v}_{-j}}(F(x | \mathbf{v}_{-j}), F(v_j | \mathbf{v}_{-j})) \cdot f(x | \mathbf{v}_{-j}), \quad (3.12)$$

onde \mathbf{v}_{-j} denota o vetor d -dimensional \mathbf{v} excluindo a j -ésima componente. Note que $c_{xv_j|\mathbf{v}_{-j}}(\cdot, \cdot)$ é a densidade de uma cópula *bivariada* marginal. Por exemplo, quando $d = 3$,

$$f(x_1|x_2, x_3) = c_{13|2}(F(x_1|x_2), F(x_3|x_2)) \cdot f(x_1|x_2)$$

e

$$f(x_2|x_3) = c_{23}(F(x_2), F(x_3)) \cdot f(x_2).$$

Expressando todas as densidades condicionais de (3.11) através de (3.12), derivamos a decomposição para $f(x_1, \dots, x_d)$ que consiste somente de distribuições marginais univariadas e cópulas bivariadas. Assim, obtemos a decomposição chamada de *pair-copula* para a densidade da cópula d -dimensional $c_{1\dots d}$, uma fatorização da cópula d -dimensional baseada apenas em cópulas bivariadas. Esta é uma maneira muito natural e flexível de construir cópulas de grandes dimensões. Note que, dada uma fatorização específica, existem várias reparametrizações possíveis.

As f.d.a.s condicionais para a construção das pair-copulas são dadas por (Joe, H., 1997)

$$F(x | \mathbf{v}) = \frac{\partial C_{x,v_j|\mathbf{v}_{-j}}(F(x | \mathbf{v}_{-j}), F(v_j | \mathbf{v}_{-j}))}{\partial F(v_j | \mathbf{v}_{-j})}.$$

No caso especial (não-condicional) onde v é univariado, e x e v são uniforme padrão, temos que

$$F(x | v) = \frac{\partial C_{xv}(x, v, \Theta)}{\partial v}$$

onde Θ é o espaço paramétrico da cópula.

Para d grande, o número de construções possíveis de uma pair-copula é muito grande. Bedford, T. e Cooke, R. (2001) mostram que há 240 decomposições diferentes quando $d = 5$. Os autores introduzem uma maneira sistemática de obter essas decomposições, a qual envolve modelos gráficos chamados de *vines* regulares. Esta representação ajuda a entender as especificações condicionais feitas para a distribuição conjunta. Como casos especiais temos as vines canônicas (*hierarchical canonical vines*), C-vines, e as D-vines. Cada um desses modelos gráficos é uma maneira específica de decompor a densidade $f(x_1, \dots, x_d)$. Por exemplo, para uma D-vine, $f()$ é igual a

$$\prod_{k=1}^d f(x_k) \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{i,i+j|i+1,\dots,i+j-1}(F(x_i|x_{i+1}, \dots, x_{i+j-1}), F(x_{i+j}|x_{i+1}, \dots, x_{i+j-1})).$$

Numa D-vine, existem $d - 1$ *trees* hierárquicas com conjuntos condicionantes crescentes, e existem $d(d - 1)/2$ cópulas bivariadas. A Figura 3.2 mostra uma decomposição D-vine para $d = 6$, que consiste em 5 árvores encaixadas, onde a árvore T_j tem $7 - j$ *nodes* e $6 - j$ cópulas bivariadas. Para uma descrição detalhada veja Aas, K., Czado, C., Frigessi, A. e Bakken, H. (2009).

O mais importante: Não é necessário que todas as cópulas bivariadas envolvidas na decomposição sejam da mesma família. Este fato exatamente o que estamos procurando, já que nosso objetivo é construir, ou estimar, uma distribuição multivariada que melhor represente os dados sendo analisados, os quais podem ter, conforme já frisamos, marginais completamente diferentes, ligadas por estruturas de dependência diversas, possuindo formas de dependência linear e não lineares, incluindo dependência de cauda, ou mesmo sendo independentes.

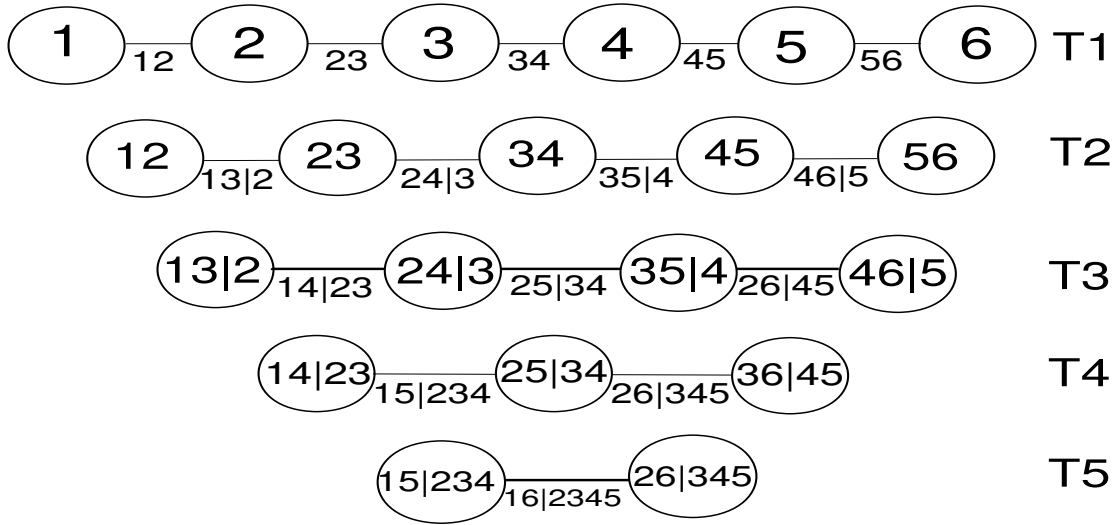


Figura 3.2: Representação gráfica de uma pair-cópula 6-dimensional tipo D-vine.

Por exemplo, poderíamos combinar os seguintes tipos de cópulas bivariadas: Gaussiana (dependência de cauda zero, elíptica); t -student (coeficientes de dependência de cauda superior e inferior iguais, elíptica); Clayton (coeficiente de dependência de cauda inferior, Archimadiana); Gumbel (coeficiente de dependência de cauda superior, Archimadiana); BB7 (coeficientes de dependência de cauda superior e inferior diferentes, Archimadiana). Veja Joe, H. (1997) para um catálogo de cópulas.

Simular tanto de uma Canônica quanto de uma D-vine é bastante simples, fácil de implementar e roda relativamente rápido. Os EMV dependem de (i) da fatorização escolhida e (ii) da escolha da família de cópulas. O algoritmo a ser implementado é simples. Para dimensões menores podemos calcular a log-verossimilhança para todas as decomposições possíveis. Para $d \geq 5$, e para a D-vine, uma decomposição específica deve ser escolhida. Uma possibilidade é escolher os pares apresentando dependência de cauda mais acentuada, e deixar esses pares determinar a decomposição a ser feita (colocá-los na *tree* 1). Para tomar esta decisão pode-se ajustar uma cópula t a todos os pares e comparar os λ_U s estimados ou os graus de liberdade. Novamente, o sucesso deste procedimento de estimação começa com bons ajustes marginais (veja Frahm, G., Junker, M. e Schmidt, R. (2004)). Os EMV podem ser utilizados nos dois passos.

Contudo, observações atípicas podem ocasionalmente ocorrer em finanças, e elas irão corromper as estimativas clássicas da estrutura de dependência. Mendes, B. V. M. e Accioly, V. B. (2013) propuseram estimar robustamente aos modelos de pair-copulas usando o *Weighted Minimum Distance (WMDE)* e os *Weighted Maximum Likelihood (WMLE)*. Estes são baseados ou numa função de peso redescendente, ou numa regra de rejeição *hard*, dicotômica, zero ou um. Notemos que a estimação robusta é efetuada no nível das cópulas bivariadas, desta maneira evitando o famoso já citado *curse of dimensionality*.

Os WMDE minimizam alguma distância ponderada (uma estatística tipo GOF) da cópula empírica.

Os WMLE resultam de um procedimento em dois passos. No primeiro passo, os *outliers* são identificados por um estimador da matriz de covariâncias com alto ponto de ruptura, e recebem peso zero. No segundo passo, os EMV dos parâmetros da cópula são calculados para o conjunto reduzido de dados. Existem vários estimadores da locação multivariada e de Σ com ponto de ruptura igual a 0.5 que podem ser empregados nesta fase preliminar. Podemos usar o estimador robusto de Stahel-Donoho (SD) (Stahel (1981) e Donoho (1982)), baseado em projeções (estão implementados no R). Este estimador afim e equivariante possui ainda função de influência limitada, e todas as boas propriedades de robustez, as quais, espera-se, se reflitam sobre as quantidades financeiras a serem calculadas, por exemplo, os pesos de ativos compondo uma carteira. Para mais detalhes sobre os métodos veja Mendes, B. V. M., Melo, E. F. L. e Nelsen, R. B. (2007).

Existem inúmeras aplicações para a modelagem com cópulas e pair-copulas em finanças. Uma aplicação diferente é quando elas são usadas para modelar a evolução dos momentos condicionais de uma série temporal. Usando cópulas podemos ir além da modelagem tradicional $\text{ARMA}(p, q)$ e $\text{GARCH}(r, s)$ (capítulos 4 e 5), as quais se baseiam somente no comportamento da f.a.c., e capturar toda a estrutura de dependência ligando observações consecutivas ou defasadas no tempo. Este tipo de dependência temporal é melhor capturada por uma decomposição tipo *Canonical vine*, veja em Mendes, B. V. M. e Aíube, C. (2011). Outras referências deste tipo de aplicação incluem Joe, H. (1997), Ibragimov (2005), e Beare, B. K. (2010). Para uma aplicação de D-vines e estimação robusta na seleção do portfolio ótimo veja Mendes, B. V. M. e Marquez, D. (2012).

Capítulo 4

Modelagem da Média Condicional

Faremos neste capítulo uma revisão dos modelos clássicos para séries temporais. Iremos a partir de agora considerar a série temporal $\{r_t\}$ dos log-retornos. Assumimos que os r_t são estacionários de 2ª ordem. Os modelos deste capítulo irão relacionar r_t com a informação disponível até o tempo $t - 1$, o conjunto $I_{t-1} = \{r_{t-1}, r_{t-2}, \dots, r_1\}$. O conjunto I_{t-1} poderá também envolver outras variáveis Y_t que descrevam o ambiente econômico e que poderiam influenciar r_t . Nesses modelos a relação de r_t com o passado será medida pela autocorrelação de *lag* k , ρ_k , também chamada de correlação serial. A teoria aqui detalhada será empregada nas diversas metodologias para o cálculo de medidas de risco que serão discutidas no Capítulo 7.

4.1 Processos estocásticos, ergodicidade e estacionariedade

Seja \mathcal{T} é um conjunto arbitrário de índices. Um processo estocástico é uma família de v.a.'s $\{r_t, t \in \mathcal{T}\}$ definidas em um espaço de probabilidade $(\Omega, \mathcal{A}, \mathcal{P})$. Em geral $\mathcal{T} = (-\infty, +\infty)$ ou mais usualmente, o conjunto dos números inteiros $\{\dots, -2, -1, 0, 1, 2, \dots\}$. Para cada t fixo temos uma v.a. r_t com a sua distribuição de probabilidade, e para todo $\omega \in \Omega$ temos $r_t(\omega)$. Assim, para cada ω temos uma realização do processo estocástico. A figura 4.1 ilustra o *ensemble* ou o conjunto de *todas* as realizações de um processo estocástico (a figura mostra algumas dessas realizações).

Como em geral observamos apenas *uma parte* de uma *única* trajetória do processo, a *série temporal* (r_1, r_2, \dots, r_T) , temos que $\mathcal{T} = \{1, \dots, T\}$ onde $T > 0$ é o tamanho da série observada. Assim temos a sequência de v.a.'s r_1, r_2, \dots, r_T e podemos descrever o processo estocástico $\{r_t, t \in \mathcal{T}\}$ através da distribuição conjunta T -dimensional de (r_1, \dots, r_T) . Isto equivale a dizer que conhecemos todas as distribuições finito-dimensionais

$$F(a_1, \dots, a_m; t_1, \dots, t_m) = P(r_{t_1} \leq a_1, \dots, r_{t_m} \leq a_m), \quad (4.1)$$

onde t_1, t_2, \dots, t_m são elementos quaisquer de \mathcal{T} , $1 \leq m \leq T$. Notemos que conhecer (4.1) significa conhecer todas as distribuições unidimensionais, das variáveis $r_t, t \in \{1, \dots, T\}$; todas as bidimensionais $(r_{t_1}, r_{t_2}), t_1, t_2 \in \{1, \dots, T\}$; etc.

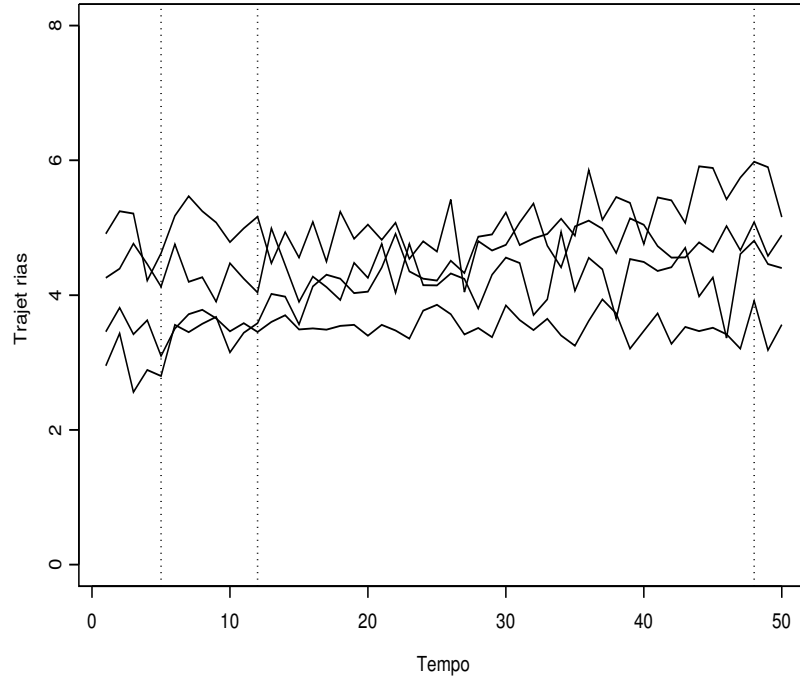


Figura 4.1: Algumas trajetórias de um processo estocástico.

Quando (r_1, \dots, r_T) tem distribuição normal T -variada é fácil obter (4.1) para todo m , pois bastam as esperanças e a matriz de covariâncias para caracterizar por completo a distribuição conjunta. Contudo, séries financeiras não são (em geral) Normais. Quando o processo $\{r_t\}$ pode ser descrito no tempo t como uma combinação linear de valores passados r_{t-1}, r_{t-2}, \dots do processo, e possivelmente também do valor corrente (tempo t) e passados de um outro processo, temos um *processo linear*. Neste caso também os dois primeiros momentos são capazes de capturar as propriedades principais do processo (veremos em 4.2 o teorema da decomposição de Wold).

Ao contrário da inferência clássica onde em geral sabemos escrever a densidade conjunta de (r_1, \dots, r_T) (onde o caso mais simples é o de v.a.'s independentes e idênticamente distribuídas, i.i.d.), aqui pouco sabemos. Poderíamos simplificar e tentar caracterizar pelo menos os T primeiros momentos, $E[r_t]$, os T segundos momentos $E[r_t^2]$, e as $\frac{T(T-1)}{2}$ covariâncias, $cov(r_i, r_j)$, $i < j$.

Contudo, tanto no caso Normal como no caso de um processo linear, temos $T + T + \frac{T(T-1)}{2}$ parâmetros desconhecidos e apenas uma realização de tamanho T do processo. Não seremos capazes de inferir os valores de todos esses parâmetros. Precisamos então reduzir o número de parâmetros desconhecidos, ou fazer outras suposições simplificadoras, obtendo algumas classes importantes de processos estocásticos, o que faremos na Seção 4.2.

Um processo é dito ser *estritamente estacionário* se a distribuição conjunta para qualquer conjunto de tempos t_1, t_2, \dots, t_m é a mesma distribuição conjunta das variáveis nos tempos $t_1+k, t_2+k, \dots, t_m+k$, para todo k , e onde m é um inteiro positivo qualquer, e (t_1, \dots, t_m) é uma coleção de m inteiros positivos. Isto é equivalente a dizer que a distribuição conjunta é invariante a um deslocamento arbitrário no tempo, uma condição muito difícil de se verificar empiricamente. Por exemplo, quando $m = 1$ a estacionariedade estrita implica que todas as v.a.'s r_t têm a mesma distribuição, e portanto mesmas médias e variâncias.

Um processo é dito ser *fracamente estacionário* ou *estacionário de segunda ordem* se tanto as distribuições univariadas assim como as bivariadas são invariantes no tempo. Isto implica que tanto a média de r_t quanto a covariância entre r_t e r_{t+k} , para qualquer inteiro k , são invariantes no tempo. Consequentemente, se uma série $\{r_t\}$ é fracamente estacionária temos que

$$E[r_1] = E[r_2] = \dots = E[r_T] \equiv \mu \text{ e } \text{var}(r_1) = \text{var}(r_2) = \dots = \text{var}(r_T) \equiv \gamma_0,$$

e que as distribuições bidimensionais dependem apenas da diferença de tempo entre as duas componentes, isto é, do *lag*. Para todo t , a autocovariância de *lag* k , denotada por γ_k é $\text{cov}(r_{t-k}, r_t) = \text{cov}(r_t, r_{t+k}) = E[r_t r_{t-k}] - E[r_t]E[r_{t-k}]$. Fazendo t igual a k (ou $-k$) temos que $\gamma_k = \text{cov}(r_0, r_k) = \text{cov}(r_{-k}, r_0)$, e portanto depende apenas das diferenças dos tempos, ou do *lag* k , k um inteiro qualquer. Veja Morettin, P. A. e Toloi, C. M. C. (2006) algumas propriedades de função de autocorrelação.

Assim, se uma série é estacionária de segunda ordem, o gráfico de seus valores no tempo devem flutuar em torno de um nível constante com variação constante. Na definição de estacionariedade fraca está implícito que assumimos serem os dois primeiros momentos de r_t finitos. Notemos que uma série estritamente estacionária com os dois primeiros momentos finitos é também fracamente estacionária. A recíproca não é verdadeira. Entretanto se uma série é fracamente estacionária e (r_1, \dots, r_T) tem distribuição T -variada normal, processo Gaussiano, então o processo é estritamente estacionário.

Observação 1: Se $\{r_t\}$ é estacionária de 2ª ordem então $\{g(r_t)\}$ também é estacionária de 2ª ordem, para qualquer função $g(\cdot)$.

Observação 2: Formas de *não-estacionariedade* podem ser sazonalidades, mudanças estruturais na série, etc.

Observação 3: A partir daqui, sempre que falarmos em série estacionária estaremos nos referindo a uma série fracamente estacionária.

A f.a.c., juntamente com as médias e as variâncias (não condicionais), nos dão uma *visão geral do processo*. A f.a.c. nos mostra como uma observação está linearmente relacionada, no tempo, com os valores anteriores da variável, e nos dá uma idéia da extensão e do grau da memória do processo. É uma ferramenta amplamente utilizada para a estimação dos parâmetros, testes de adequacidade do modelo, etc. Vimos na Seção 1.2 os estimadores amostrais $\hat{\rho}_k$ e $\hat{\gamma}_k$ e as versões amostrais da f.a.c. e da função de autocovariância (f.a.cov.), além do teste de Ljung-Box baseado nos $\hat{\rho}_k^2$.

Seja $\{r_t^{(i)}, t = 1, \dots, T\}$ uma das I realizações do processo (I elementos do *ensemble*). Pelo teorema central do limite temos que para todo t o limite em probabilidade quando $I \rightarrow \infty$ de $\frac{1}{I} \sum_{i=1}^I r_t^{(i)}$ é dado por $E(r_t)$, ou seja, teríamos que ter várias realizações do processo para se conhecer $E(r_t) \forall t$. Não temos isto. Precisamos então de assumir a propriedade de ergodicidade.

Consideremos a média no tempo, $\frac{1}{T} \sum_{t=1}^T r_t^{(1)} = \bar{r}$ para a realização (1), isto é, para a única

realização que temos. A *ergodicidade* garante que $\bar{r} \xrightarrow{p} E[r_t]$ para um processo estacionário. Assim, se $\bar{r} \xrightarrow{p} E[r_t]$ em um processo estacionário de 2ª ordem, este processo é dito ergódico para a média. Pode-se mostrar que se

$$\sum_{k=0}^{\infty} |\gamma_k| < \infty$$

então $\{r_t\}_{-\infty}^{+\infty}$ é ergódico para a média.

Similarmente, um processo covariância-estacionário é ergódico para os segundos momentos se

$$\hat{\gamma}_j \equiv \frac{1}{(T-j)} \sum_{t=j+1}^T (r_t - \mu)(r_{t-j} - \mu) \xrightarrow{p} \gamma_j \quad \forall j$$

Por exemplo, um processo estacionário Gaussiano é ergódico para os momentos de ordem 2. Ergodicidade garante que quando $T \rightarrow \infty$, os momentos amostrais se aproximam dos momentos populacionais.

Voltaremos a falar em (não) estacionariedade nas seções 4.5 e 4.6.

4.2 Exemplos de processos estocásticos

Considere a *sequência aleatória* $\{r_t, t = 1, 2, \dots\}$, isto é, uma sequência de v.a.'s definidas no mesmo espaço amostral Ω . Para todo $t \geq 1$ podemos escrever

$$P(r_1 = x_1, \dots, r_t = x_t) = P(r_1 = x_1)P(r_2 = x_2 | r_1 = x_1) \cdots P(r_t = x_t | r_1 = x_1, \dots, r_{t-1} = x_{t-1}) \quad (4.2)$$

onde os x_j 's representam os estados do processo e o espaço dos estados pode ser \mathfrak{R} . Vejamos alguns casos particulares da sequência aleatória.

Sequência de v.a.'s mutuamente(coletivamente) independentes. Neste caso, $\{r_t, t \geq 1\}$ é uma sequência de v.a.'s mutuamente independentes (i.). Neste caso, as probabilidades em (4.2) simplificam:

$$P(r_1 = x_1, \dots, r_t = x_t) = P(r_1 = x_1)P(r_2 = x_2) \cdots P(r_t = x_t) \quad (4.3)$$

Processo puramente aleatório. Ao processo acima incluímos a restrição de serem as v.a.'s idênticamente distribuídas (i.d.), e teremos um processo de v.a.'s i.i.d. Neste caso o processo é estacionário com $E[r_t] = \mu$ e $\gamma_0 = \text{var}(r_t) = \sigma^2 \forall t$, com $\sigma^2 < \infty$, e $\gamma_k = 0 \forall k \geq 1$. Portanto $\rho_0 = 1$ e $\rho_k = 0 \forall k \geq 1$. Se $\{r_t\}$ é um processo puramente aleatório, o denotamos por

$$r_t \sim \text{i.i.d.}(\mu, \sigma^2).$$

A densidade conjunta deste processo é um caso particular de (4.3), onde todas as densidades univariadas são iguais.

Ruído Branco. As v.a.'s são i.d.'s não correlacionadas e com variância finita, tendo portanto a mesma f.a.c. de um processo i.i.d. Denotamos por

$$r_t \sim \text{RB}(\mu, \sigma^2),$$

cuja densidade conjunta não necessariamente fatora nas marginais.

A Figura 4.2 mostra à esquerda um processo puramente aleatório, onde r_t tem distribuição $t\text{-student}(\nu)$ com $\nu=3$ e portanto tem variância igual a $\frac{\nu}{\nu-2}$. À direita temos um ruído branco onde a distribuição não condicional tem caudas mais pesadas que a normal e variância também igual a 3.

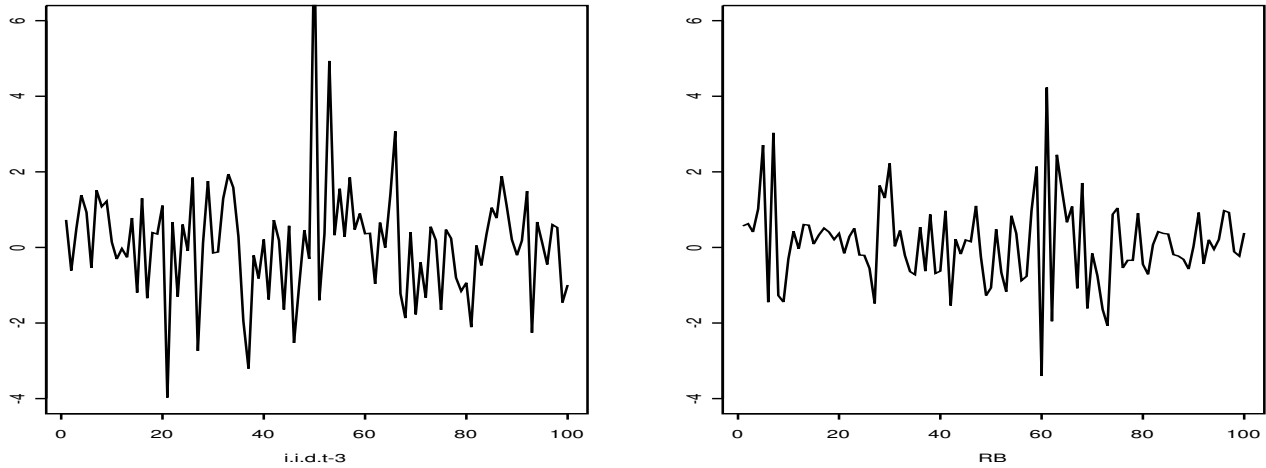


Figura 4.2: Dois processos estocásticos com variância igual a 3. À esquerda um puramente aleatório; à direita um ruído branco.

Observação: Um $RB(\mu, \sigma^2)$ Gaussiano, isto é, um RB onde a distribuição de cada r_t é $Normal(\mu, \sigma^2)$, somente será um processo i.i.d. se a cópula de (r_1, \dots, r_T) for a cópula Gaussiana com coeficientes de correlação iguais a zero ou for a cópula Produto.

Passeio Aleatório. Considere a sequência aleatória $\{\varepsilon_t, t \geq 1\}$ i.i.d. $(\mu_\varepsilon, \sigma_\varepsilon^2)$. O passeio aleatório (PA) é definido como o processo $\{X_t\}$ tal que

$$X_t = X_{t-1} + \varepsilon_t \quad ,$$

com $X_0 \in \Re$ um valor inicial do processo. Note que $E[X_t] = t\mu_\varepsilon$ e $var(X_t) = t\sigma_\varepsilon^2$. Pode-se mostrar que a covariância entre X_i e X_j é igual a $\sigma_\varepsilon^2 \min(i, j)$, e portanto aumenta com o tempo. É um processo não-estacionário. Mas como $X_t - X_{t-1} = \varepsilon_t$, dizemos que este processo tem “incrementos” não correlacionados. Tem bastante importância em finanças já que o processo dos log-preços $\{p_t\}$ tem características similares a um PA (neste caso p_0 seria o preço inicial IPO, *initial public offering*). Séries de retornos financeiros em geral apresentam média pequena, positiva ou negativa. Isto implica que o modelo para os log-preços seria um PA com *drift*:

$$p_t = \mu + p_{t-1} + \varepsilon_t.$$

O termo constante $\mu = E(p_t - p_{t-1})$, representa a tendência no tempo do log-preço e é chamado de *drift* do processo. A suposição de incrementos i.i.d. implica em variações de preço com esperança constante e retornos não correlacionados, mas também que qualquer função não linear dos retornos sejam também não correlacionadas (veja Campbell, J. Y., Lo, A. W. e MacKinley, A. C (1997)). Figura 4.3 mostra um passeio aleatório a partir de um processo i.i.d. com distribuição t-student(3).

O modelo básico em finanças assume que o preço de um ativo no tempo t , P_t , segue um movimento Browniano, solução de uma equação diferencial estocástica de Itô. Este modelo serve de base para vários outros, como por exemplo, o modelo de precificação de Black-Scholes. Uma consequência deste modelo é que os log-retornos seriam normais independentes e idênticamente distribuídos. Mais especificamente, o modelo prescreve que $\log(P_t) = ct + \sigma W_t$, $t \geq 0$, onde $\sigma, c > 0$ e W é um movimento Browniano. Assim, os log-retornos seriam Gaussianos: $r_t = c + \sigma(W_t - W_{t-1})$. Esta suposição é conveniente sob o ponto de vista matemático mas não é realista.

Veja mais sobre PA na seção 4.6.

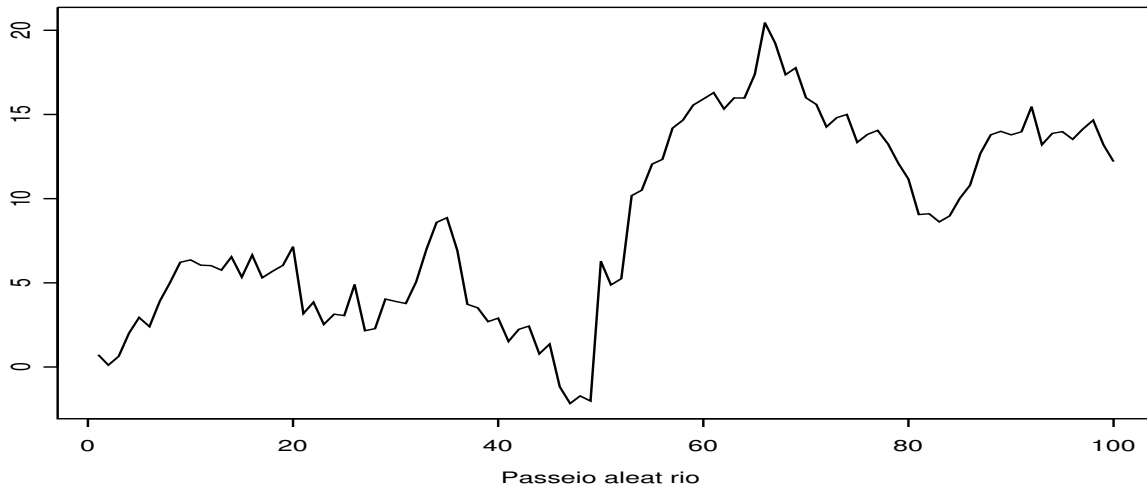


Figura 4.3: Um passeio aleatório a partir de um processo i.i.d. com distribuição t-student(3).

Martingales. Seja $\{X_t, t \in \mathcal{I}\}$, com $\mathcal{I} \subset \mathbb{R}$, um processo estacionário, isto é, para cada $t \in \mathcal{I}$, X_t é uma v.a. definida em um espaço de probabilidade (Ω, \mathcal{F}, P) . Seja $\{\mathcal{F}_t\}$ uma filtração genérica que representa os acréscimos de informação ao longo do tempo, e portanto, a sigma-álgebra $\mathcal{F}_t = \sigma(\{X_s : s \leq t\})$, representa a informação disponível no tempo t . Seja $\{X_t\}$ um processo \mathcal{F}_t -mensurável e integrável. $\{X_t\}$ é um \mathcal{F}_t -martingale se $E[X_s | \mathcal{F}_t] = X_t$ para todo $0 \leq t \leq s$, isto é, o valor corrente X_t é o melhor preditor (com função de perda quadrática) de um valor futuro X_s . Veja o Apêndice AP1.1 do Capítulo 1.

Numa sequência *martingale-difference*, o valor esperado do próximo valor, dada a informação cor-

rente, é sempre zero. Notemos que esta é uma propriedade que retornos financeiros podem apresentar. Este modelo é geralmente utilizado para modelar os ganhos em *rounds* sucessivos de um jogo honesto.

Assim, uma série temporal $\{Y_t\}_{t \in \mathbf{Z}}$ é uma sequência *martingale-difference* em relação a uma filtração $\{\mathcal{F}_t\}_{t \in \mathbf{Z}}$ se $E|Y_t| < \infty$, se Y_t é \mathcal{F}_t -mensurável e $E[Y_t|\mathcal{F}_{t-1}] = 0$, $\forall t \in \mathbf{Z}$. Observemos que a média não condicional também é zero: $E[Y_t] = E[E[Y_t|\mathcal{F}_{t-1}]] = 0$, $\forall t \in \mathbf{Z}$. Pode-se mostrar que se $E[Y_t^2] < \infty \forall t$, então todas as autocovariâncias $\gamma(t, s) = E[Y_t Y_s] = 0$. Notemos que se a variância (finita) for constante $\forall t$, um processo *martingale-difference* com média e covariâncias zero será um ruído branco.

Pela sua importância, os processos ARMA(p, q) ou mais geralmente ARFIMA(p, d, q) serão dados na seção 4.4.

4.3 Processos lineares

O teorema da decomposição de Wold (Wold, H., 1938) estabelece que todo processo estocástico fracamente estacionário com média zero, $(r_t - \mu)$, pode ser descrito como uma combinação linear de uma sequência de v.a.'s independentes e idênticamente distribuídas com média zero. Este filtro linear deve ser passado após a retirada de qualquer componente determinístico tais como variações periódicas, ciclos, tendências, etc. Este teorema é a base dos modelos clássicos para séries temporais.

A decomposição de uma série estacionária numa combinação linear (ou filtro linear) pode ser representada como

$$r_t - \mu = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots = \sum_{j=0}^{\infty} \psi_j a_{t-j} \quad \text{com } \psi_0 = 1, \quad (4.4)$$

onde os a_t são as inovações, uma sequência de v.a.'s independentes, com a mesma distribuição, e com

$$E[a_t] = 0, \quad \text{var}(a_t) = E[a_t^2] = \sigma_a^2 < \infty$$

e

$$\text{cov}(a_t, a_{t-k}) = E[a_t a_{t-k}] = 0, \quad \forall k \neq 0.$$

Notemos que Tsay, R. S. (2005) assume $a_t \sim \text{i.i.d.}(0, \sigma_a^2)$ e Morettin, P. A. (2008) assume $a_t \sim \text{RB}(0, \sigma_a^2)$.

Nem todas as séries financeiras são lineares. Estudaremos alguns modelos não-lineares no Capítulo 5. Os coeficientes do filtro linear (talvez um número infinito deles), os pesos ψ_j são chamados de *impulse response* devido ao fato de que $\psi_j = \frac{\partial r_{t+j}}{\partial a_t}$, $j = 1, 2, \dots$. O gráfico dos pesos ψ_j versus j dá origem à função IRF, *Impulse Response Function*, veja em Tsay, R. S. (2005).

Considere novamente (4.4). Calculemos os momentos não condicionais do processo linear fracamente estacionário. Temos:

$$E[r_t] \equiv \mu$$

$$\gamma_0 \equiv \text{var}(r_t) = E[(r_t - \mu)^2] = E[(a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots)^2].$$

Como as inovações a_t são independentes, os momentos cruzados $E[a_t a_{t-k}]$ desaparecem, e temos

$$\gamma_0 = E[a_t^2] + \psi_1^2 E[a_{t-1}^2] + \psi_2^2 E[a_{t-2}^2] + \dots$$

$$\gamma_0 = \sigma_a^2 + \psi_1^2 \sigma_a^2 + \psi_2^2 \sigma_a^2 + \dots$$

$$\gamma_0 = \text{var}(r_t) = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j^2.$$

A covariância de *lag* k do processo r_t ou do processo $r_t - \mu$ é

$$\begin{aligned} \gamma_k &= E[(r_t - \mu)(r_{t-k} - \mu)] \\ &= E[(a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots)(a_{t-k} + \psi_1 a_{t-k-1} + \psi_2 a_{t-k-2} + \dots)] \\ &= E[(a_t + \psi_1 a_{t-1} + \dots + \psi_k a_{t-k} + \dots)(a_{t-k} + \psi_1 a_{t-k-1} + \dots)] \\ &= E[(\psi_k a_{t-k}^2 + \psi_1 \psi_{k-1} a_{t-k-1}^2 + \dots)] \\ &= \sum_{j=0}^{\infty} \psi_j \psi_{j+k} E[a_{t-k-j}^2] \end{aligned}$$

Isto é,

$$\gamma_k = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k},$$

o que implica

$$\rho_k = \frac{\sum_{j=0}^{\infty} \psi_j \psi_{j+k}}{\sum_{j=0}^{\infty} \psi_j^2}.$$

Vemos que o numerador da fórmula das autocorrelações é uma soma infinita de produtos dos pesos, e que o denominador é uma soma infinita dos quadrados desses pesos.

Para que se tenha uma série estacionária e ergódica, temos que assumir que a série converge, isto é, devemos ter $\lim_{j \rightarrow \infty} \psi_j \rightarrow 0$ e $\sum_{j=0}^{\infty} \psi_j < \infty$. Uma suposição usualmente feita para garantir convergência é a de que

$$\sum_{j=0}^{\infty} j |\psi_j| < \infty,$$

chamada de *1-summability*.

Assim, o padrão das autocorrelações em todo o processo estacionário ergódico $\{r_t\}$ é determinado pelo conjunto dos pesos $\{\psi_j\}$, da sua representação como um processo linear segundo o teorema da decomposição de Wold.

A representação de Wold usando o operador retroativo B é

$$r_t = \mu + \Psi(B)a_t$$

$$\Psi(B) = \sum_{k=0}^{\infty} \psi_k B^k, \quad \psi_0 = 1.$$

Observações: $B^s a_t = a_{t-s}$; $\Psi(1) = \sum_{k=0}^{\infty} \psi_k$, a soma dos coeficientes.

Podemos aproximar o polinômio $\Psi(B)$ pela razão de 2 polinômios finitos, $\Psi(B) = \frac{\Theta(B)}{\Phi(B)}$, e obter modelos para séries temporais estacionárias, como por exemplo, um modelo ARMA. Um modelo ARMA(p, q), que será visto na Seção 4.4, é dado por

$$r_t = \mu + \Psi(B)a_t$$

$$r_t - \mu = \frac{\Theta(B)}{\Phi(B)}a_t$$

$$\Phi(B)(r_t - \mu) = \Theta(B)a_t$$

$$r_t - \mu = \phi_1(r_{t-1} - \mu) + \cdots + \phi_p(r_{t-p} - \mu) + a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q}$$

com $\{a_t\}$ uma sequência de v.a.'s não correlacionadas com média zero e variância σ_a^2 constante.

4.4 Modelagem AR(F)IMA

A evolução temporal de uma série pode ser descrita através de processos (integrados) autorregressivos e de média móvel, ARIMA(p, d, q). A classe dos modelos ARIMA, popularizada por Box, G. E. P., Jenkins, G. M. e Reinsel, G. (1994), tem como casos especiais os modelos AR(p), MA(q) e ARMA(p, q).

Considere o processo $\{p_t\}$, onde p_t representa o logaritmo do preço P_t de um ativo ou carteira no dia (ou qualquer unidade de tempo) t . O processo $\{p_t\}$ segue um modelo ARIMA(p, d, q), d inteiro, se $\Delta^d p_t \equiv (1 - B)^d p_t = r_t$ seguir um ARMA(p, q), isto é,

$$r_t = \phi_0 + \phi_1 r_{t-1} + \cdots + \phi_p r_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}, \quad (4.5)$$

onde $\phi_0 = \mu(1 - \phi_1 - \cdots - \phi_p)$, e onde $a_t \sim \text{RB}(0, \sigma_a^2)$. Supondo $\mu = 0$, temos $\Phi(B)\Delta^d p_t = \Theta(B)a_t$, onde B é o operador retroativo. A diferenciação Δ^d é necessária para estacionarizar a série. Se após d diferenças a série resultante for estacionária dizemos que a série original $\{p_t\}$ é integrada de ordem d , uma série I(d), e a série final, $\{r_t\}$ é uma série I(0). Em finanças, em geral basta uma diferenciação, isto é, $d = 1$ para séries de preços diários. Existem várias estatísticas para se testar estacionariedade. Por exemplo, o teste KPSS (Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. e Shin, Y. (1992)) visto no Capítulo 1, cuja hipótese alternativa é “existência de raiz unitária ou memória longa”, ou o teste ADF, de Dickey, D. A. e Fuller, W. A. (1979), derivado a partir de modelos autorregressivos. Aplicamos os dois testes nas três séries de retornos aqui utilizadas e todos os testes confirmaram a hipótese nula de estacionariedade.

Para facilitar a identificação dos submodelos daremos a seguir uma revisão das propriedades de cada um deles.

4.4.1 Processos Autorregressivos AR(p)

Um processo autorregressivo de ordem p assume que

$$r_t = \phi_0 + \phi_1 r_{t-1} + \cdots + \phi_p r_{t-p} + a_t,$$

onde $\{a_t\}$ é uma série de RB($0, \sigma_a^2$).

Processos AR(1)

Considere o caso particular $p = 1$, $r_t = \phi_0 + \phi_1 r_{t-1} + a_t$, onde $\{a_t\} \sim \text{RB}(0, \sigma_a^2)$, e calculemos as esperanças e variâncias condicionais e não condicionais.

$$E[r_t | r_{t-1}] = \phi_0 + \phi_1 r_{t-1} \quad e \quad \text{var}(r_t | r_{t-1}) = \sigma_a^2.$$

A esperança não condicional pode ser calculada a partir de

$$E[r_t] = E[E[r_t | r_{t-1}]] = E[\phi_0 + \phi_1 r_{t-1}] = \phi_0 + \phi_1 E[r_{t-1}],$$

e como o processo é estacionário, $E[r_t] = E[r_{t-1}]$, e obtemos

$$\mu \equiv E[r_t] = \frac{\phi_0}{1 - \phi_1}.$$

Notamos então que (i) $\phi_1 \neq 1$ para μ existir; (ii) μ será zero se e somente se (sss) $\phi_0 = 0$.

Para achar a variância não condicional de r_t pode-se usar a fórmula

$$\text{var}(r_t) = E[\text{var}(r_t | r_{t-1}, \dots)] + \text{var}(E[r_t | r_{t-1}, \dots]),$$

ou, equivalentemente, a partir de

$$r_t - \mu = \phi_1(r_{t-1} - \mu) + a_t,$$

obter

$$\text{var}(r_t - \mu) = \text{var}(r_t) = \phi_1^2 \text{var}(r_{t-1}) + \sigma_a^2,$$

já que a_t e a_{t-1} são não correlacionados. Como, por estacionariedade, $\text{var}(r_{t-1}) = \text{var}(r_t)$, temos

$$\gamma_0 \equiv \text{var}(r_t) = \frac{\sigma_a^2}{1 - \phi_1^2},$$

desde que $\phi_1^2 < 1$. Assim, um AR(1) fracamente estacionário tem $|\phi_1| < 1$. Por outro lado, pode-se mostrar que $|\phi_1| < 1$ implica que o AR(1) é fracamente estacionário. Para ver isto reescrevemos o modelo $r_t = \phi_1 r_{t-1} + a_t$ através de substituições sucessivas até obter

$$r_t = \sum_{j=0}^s \phi_1^j a_{t-j} + \phi_1^{s+1} r_{t-s-1}.$$

Sendo r_t estacionária com variância finita γ_0 obtemos

$$\begin{aligned}
E[(r_t - \sum_{j=0}^s \phi_1^j a_{t-j})^2] &= \phi_1^{2s+2} E(r_{t-s-1}^2) \\
&= \phi_1^{2(s+1)} \gamma_0.
\end{aligned}$$

Agora, se $|\phi_1| < 1$, o termo $\phi_1^{2(s+1)} \rightarrow 0$, quando $s \rightarrow \infty$, e temos

$$E[(r_t - \sum_{j=0}^s \phi_1^j a_{t-j})^2] \rightarrow 0,$$

o que significa que, quando $s \rightarrow \infty$

$$r_t \xrightarrow{MQ} \sum_{j=0}^{\infty} \phi_1^j a_{t-j},$$

isto é, converge em média quadrática e temos

$$r_t = \sum_{j=0}^{\infty} \phi_1^j a_{t-j}.$$

Assim, escrevendo r_t como uma combinação linear de um número infinito de inovações, e sabendo que $-1 < \phi_1 < +1$, podemos provar que $E[r_t] < \infty$ e $var(r_t) < \infty$, sendo $\gamma_0 = var(\sum_{j=0}^{\infty} \phi_1^j a_{t-j}) = \sigma_a^2 \sum_{j=0}^{\infty} \phi_1^{2j} = \frac{\sigma_a^2}{1-\phi_1^2}$. Além disto, usando a desigualdade de Cauchy-Schwartz podemos mostrar que as autocovariâncias são finitas. Em resumo, um $AR(1)$ é fracamente estacionário se e somente se $|\phi_1| < 1$.

A Figura 4.4 mostra as trajetórias de dois processos $AR(1)$ com $\phi_0 = 0$, tendo o da esquerda $\phi_1 = 0.05$, e o da direita $\phi_1 = 0.95$, ambos com a_t normal com $\sigma_a^2 = 0.09$ e $T = 100$. Note que as variâncias não condicionais dos processos são, respectivamente, 0.0902 e 0.9231.

Para obter a função de autocorrelação de um processo $AR(1)$ estacionário com $\mu = \phi_0/(1 - \phi_1)$, escrevemos $\phi_0 = \mu(1 - \phi_1)$ e então

$$r_t - \mu = \phi_1(r_{t-1} - \mu) + a_t. \quad (4.6)$$

Queremos achar $\gamma_k \equiv cov((r_t - \mu), (r_{t-k} - \mu)) = E[(r_t - \mu)(r_{t-k} - \mu)]$. Multiplicamos (4.6) por $(r_{t-k} - \mu)$, tomamos esperanças, e como $E[(r_{t-k} - \mu)a_t] = 0 \forall k \geq 1$, obtemos

$$\gamma_k \equiv E[(r_t - \mu)(r_{t-k} - \mu)] = E[\phi_1(r_{t-1} - \mu)(r_{t-k} - \mu)] = \phi_1 \gamma_{k-1}.$$

Então $\gamma_k = \phi_1^k \gamma_0$, com $\gamma_0 = \frac{\sigma_a^2}{1-\phi_1^2}$. Assim temos as expressões para os ρ_k :

$$\rho_0 = 1 \quad \rho_k = \frac{\phi_1^k \gamma_{k-1}}{\gamma_0}, \quad k \geq 1.$$

Isto é, $\rho_k = \phi_1^k \rho_{k-1}$ e obtemos

$$\rho_k = \phi_1^k \quad k \geq 0.$$

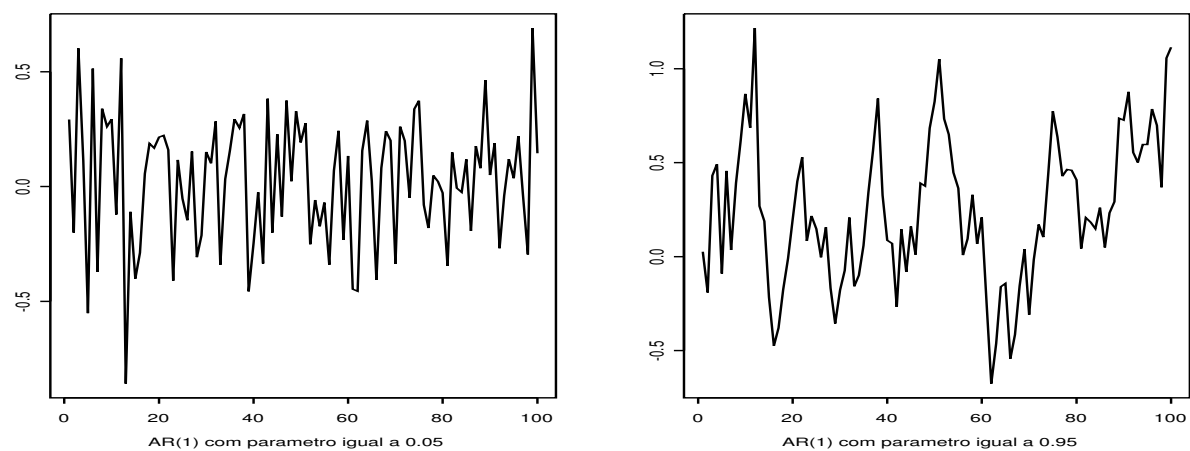


Figura 4.4: Trajetórias de dois processos $AR(1)$ com $\phi_0 = 0$, $\phi_1 = 0.05$, $\phi_1 = 0.95$, a_t normal com $\sigma^2 = 0.09$ e $T=100$.

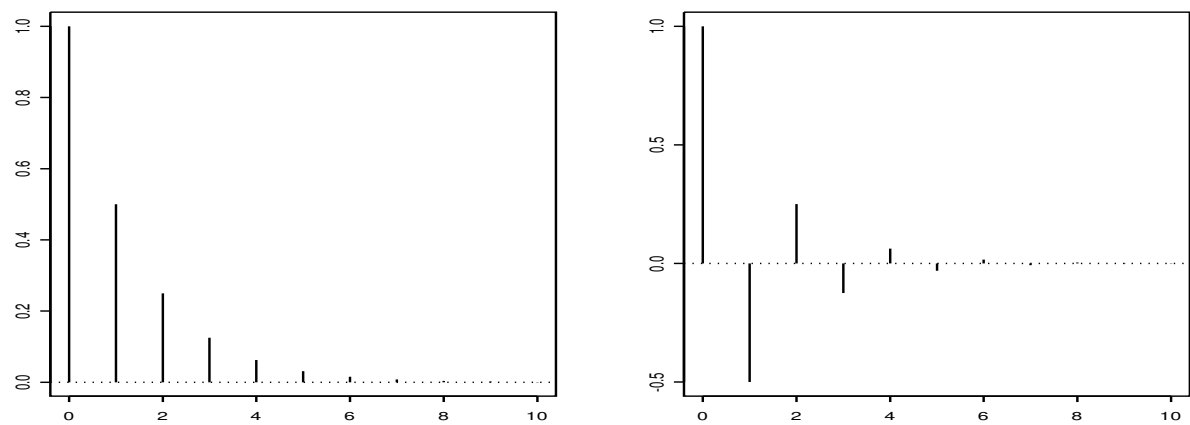


Figura 4.5: As f.a.c.'s de processos $AR(1)$ com $\phi_1 = 0.5$ do lado esquerdo e $\phi_1 = -0.5$ do lado direito.

Vemos que a f.a.c. de um AR(1) decai exponencialmente à taxa ϕ_1 e com valor inicial $\rho_0 = 1$. A Figura 4.5 mostra o decaimento da f.a.c. de um AR(1) com $\phi_1 = 0.5$ do lado esquerdo e $\phi_1 = -0.5$ do lado direito.

A estimação do modelo AR(1) é em geral feita por máxima verossimilhança condicional, mas também podem ser utilizados os métodos dos momentos ou de mínimos quadrados.

Para se obter as estimativas or máxima verossimilhança (EMV) é preciso supor uma distribuição para a inovação a_t . Supondo a_t com distribuição normal e *condicional* ao valor inicial da série, r_0 , a função de verossimilhança é dada por

$$L((\phi_0, \phi_1, \sigma_a^2) | (r_1, \dots, r_T), r_0) = (2\pi\sigma_a^2)^{-T/2} \exp\left\{\frac{-1}{2\sigma_a^2} \sum_{t=1}^T (r_t - \phi_0 - \phi_1 r_{t-1})^2\right\}.$$

Os EMV são obtidos a partir da maximização da log-verossimilhança \mathcal{L} , a qual fica proporcional a

$$\mathcal{L}((\phi_0, \phi_1, \sigma_a^2) | (r_1, \dots, r_T), r_0) \sim (-T/2) \log(\sigma_a^2) - \left(\frac{1}{2\sigma_a^2}\right) \sum_{t=1}^T (r_t - \phi_0 - \phi_1 r_{t-1})^2.$$

Notemos que maximizar \mathcal{L} com σ_a^2 fixo equivale a obter os estimadores de mínimos quadrados (MQ).

Uma outra possibilidade é estimar os parâmetros a partir de um modelo de regressão, e neste caso utilizamos apenas os valores observados, sem necessidade de supor um valor r_0 . Os resíduos serão dados por $\hat{a}_t = r_t - \hat{r}_t$, onde $\hat{r}_t = \hat{\phi}_0 + \hat{\phi}_1 r_{t-1}$, e um estimador para a variância das inovações pode ser obtido a partir de $\hat{\sigma}_a^2 = \frac{\sum_{t=2}^T \hat{a}_t^2}{T-3}$. A constante do denominador segue do fato de que temos $T-1$ termos no somatório do numerador e 2 parâmetros a estimar.

Alternativamente, pode-se obter estimadores do modelo AR(1) a partir das equações diferença de Yule-Walker

$$\rho_0 \phi_1 = \rho_1,$$

que não apresentam dificuldades computacionais (Brockwell e Davis, 1991).

Por exemplo, a estimação por MV do modelo AR(1) para a série dos retornos brasileiros resultou em:

```
Method: Maximum Likelihood
Model : 1 0 0
Coefficients: AR : 0.05367
Variance-Covariance Matrix:
      ar(1)
ar(1) 0.0007353391
Optimizer has converged
erro padrao = 0.02711714
estatistica t = 1.9792
$gof$p.value (Portmanteau teste)
0.0178 0.0407 0.0869 0.0017 0.0038
```

A Figura 4.6 mostra as análises gráficas deste ajuste AR(1) aos retornos brasileiros. Observe o gráfico dos resíduos no tempo e veja que as características da série associadas ao seu segundo momento, os *clusters* de volatilidade, ainda estão lá. É um tipo de dependência não linear, não capturada pelo modelo linear ajustado para a média.

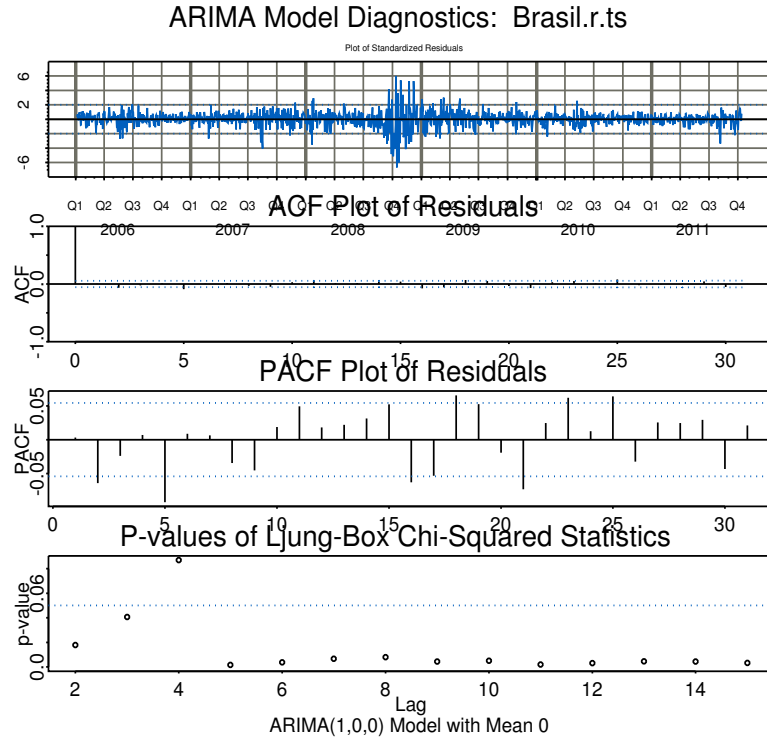


Figura 4.6: Análise gráfica do ajuste de um AR(1) aos retornos do índice brasileiro.

Processos AR(2)

Seja

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + a_t.$$

Assim como fizemos no modelo AR(1), obtemos agora

$$E(r_t) = \frac{\phi_0}{1 - \phi_1 - \phi_2} = \mu \quad \text{onde} \quad \phi_1 + \phi_2 \neq 1$$

Usando $\mu(1 - \phi_1 - \phi_2) = \phi_0$, reescrevemos o AR(2) como:

$$\begin{aligned} r_t &= \mu(1 - \phi_1 - \phi_2) + \phi_1 r_{t-1} + \phi_2 r_{t-2} + a_t \\ r_t - \mu &= -\phi_1 \mu - \phi_2 \mu + \phi_1 r_{t-1} + \phi_2 r_{t-2} + a_t \\ r_t - \mu &= \phi_1 (r_{t-1} - \mu) + \phi_2 (r_{t-2} - \mu) + a_t \end{aligned} \tag{4.7}$$

Multiplicando (4.7) por $(r_{t-k} - \mu)$, tomando esperanças, e sabendo que $E[(r_{t-k} - \mu)a_t] = 0$ para $k > 0$, obtemos:

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}, \quad \text{para } k > 0. \quad (4.8)$$

(4.8) é chamada de equações dos momentos de um AR(2) estacionário. Dividindo (4.8) por γ_0 obtemos

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, \quad k > 0. \quad (4.9)$$

Assim,

$$\rho_1 = \phi_1 \rho_0 + \phi_2 \rho_{-1} = \phi_1 + \phi_2 \rho_1 \quad \Rightarrow \quad \rho_1 = \frac{\phi_1}{1 - \phi_2}$$

e os outros ρ_k

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, \quad k \geq 2.$$

Da equação (4.8) notamos que a f.a.c. de um AR(2) satisfaz uma equação de diferenças de 2ª ordem:

$$(1 - \phi_1 B - \phi_2 B^2) \rho_k = 0 \quad (4.10)$$

Existe uma equação polinomial de 2ª ordem correspondente à equação de diferenças de 2ª ordem (4.10), que é:

$$x^2 - \phi_1 x - \phi_2 = 0$$

cujas raízes são $x = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2}$. Estas soluções são chamadas de raízes características do AR(2). Chamemos estas raízes de w_1 e w_2 . Se ambas w_1 e w_2 forem reais, $(1 - \phi_1 B - \phi_2 B^2)$ pode ser fatorado em $(1 - w_1 B)(1 - w_2 B)$:

$$(1 - \phi_1 B - \phi_2 B^2) r_t = a_t$$

$$(1 - w_1 B)(1 - w_2 B) r_t = a_t$$

e isto nos faz ver um AR(2) como se fosse um AR(1) operando sobre um outro AR(1). É por isto que a f.a.c., equação (4.10), fica uma mistura de dois decaimentos exponenciais. Contudo se $\phi_1^2 + 4\phi_2 < 0$, então w_1 e w_2 são raízes complexas e a f.a.c. será ondas de senos e cossenos amortecidas (em economia seriam interpretadas como os *business cycles*).

Um AR(2) é estacionário se suas duas raízes características, relacionadas à $x^2 - \phi_1 x - \phi_2 = 0$, são menores que 1. Pode-se mostrar que a condição | raízes características | < 1 implica no seguinte conjunto de restrições para ϕ_1 e ϕ_2 :

$$\phi_1 + \phi_2 < 1; \quad -\phi_1 + \phi_2 < 1; \quad -1 < \phi_2 < 1.$$

Observemos que $\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}$ e portanto as autocovariâncias seguem a mesma equação diferença de segunda ordem do AR(2). Portanto, um AR(2) é covariância-estacionário se ϕ_1 e ϕ_2 pertencem a um conjunto bem definido pelas restrições, veja Hamilton, J. D. (1994). Pode-se mostrar que $\gamma_0 = \frac{(1-\phi_2)\sigma_a^2}{(1+\phi_2)[(1-\phi_2)^2 - \phi_1^2]}$.

Um exercício interessante é gerar várias séries AR(2) a partir de valores para os parâmetros e observar o comportamento das mesmas no tempo e os gráficos de suas f.a.c.'s. Por exemplo, sugerimos simular 4 processos AR(2) com $\phi_1 = \pm 0.5$ e $\phi_2 = \pm 0.3$.

Processos AR(p)

Um processo autorregressivo de ordem p assume que

$$r_t = \phi_0 + \phi_1 r_{t-1} + \dots + \phi_p r_{t-p} + a_t, \quad (4.11)$$

onde $a_t \sim \text{RB}(0, \sigma_a^2)$. A média e a variância não condicionais de um processo AR(p) estacionário são dadas por

$$\mu \equiv E[r_t] = \frac{\phi_0}{1 - \phi_1 - \dots - \phi_p}.$$

$$\sigma^2 \equiv \gamma_0 \equiv \text{var}(r_t) = \frac{\sigma_a^2}{1 - \phi_1^2 - \dots - \phi_p^2}.$$

A esperança condicional depende dos p retornos passados e a variância condicional é σ_a^2 .

Supondo $\mu = 0$ temos

$$\phi(B)r_t = a_t$$

onde $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ é o operador retroativo de ordem p .

Pode-se mostrar que um AR(p) é estacionário e ergódico se as raízes da equação característica

$$\phi(x) = 1 - \phi_1 x - \phi_2 x^2 - \dots - \phi_p x^p = 0 \quad (4.12)$$

tiverem valor absoluto maior que 1 (fora do disco unitário). Uma condição necessária para estacionariedade é que $|\phi_1 + \dots + \phi_p| < 1$, bastante útil na prática.

Para um AR(p) estacionário, os ρ'_k s satisfazem a equação de diferenças

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)\rho_k = 0, \quad k > 0.$$

que são as equações de Yule-Walker.

O gráfico da f.a.c. será então uma mistura de decaimentos exponenciais se todas as raízes forem reais; ou uma mistura de decaimentos exponenciais com padrões de senos e cossenos se existirem raízes complexas.

Para a determinação da ordem de um modelo AR(p), usaremos a f.a.c., a função de autocorrelação parcial, f.a.c.p., e algum critério de informação. A f.a.c.p., obtida a partir das estimativas dos coeficientes dos modelos AR(p) em ordens consecutivas, $p = 1, 2, \dots$, é mais útil para identificar um processo AR(p). Os coeficientes de autocorrelação parcial $\phi_{k,k}$ medem a contribuição adicional de um modelo com ordem mais alta em relação ao anterior. Assim, se a série foi gerada a partir de um processo AR(p), a estimativa do coeficiente $\phi_{p,p}$ deve ser significativamente diferente de zero, e as estimativas de $\phi_{k,k}$, $k > p$ devem estar próximas de zero.

Para a construção da f.a.c.p., considere as regressões lineares múltiplas associadas aos modelos AR:

$$r_t = \phi_{0,1} + \phi_{1,1}r_{t-1} + \varepsilon_{1,t} \quad (1)$$

$$r_t = \phi_{0,2} + \phi_{1,2}r_{t-1} + \phi_{2,2}r_{t-2} + \varepsilon_{2,t} \quad (2)$$

$$r_t = \phi_{0,3} + \phi_{1,3}r_{t-1} + \phi_{2,3}r_{t-2} + \phi_{3,3}r_{t-3} + \varepsilon_{3,t} \quad (3)$$

$$r_t = \phi_{0,4} + \phi_{1,4}r_{t-1} + \phi_{2,4}r_{t-2} + \phi_{3,4}r_{t-3} + \phi_{4,4}r_{t-4} + \varepsilon_{4,t} \quad (4)$$

$$\vdots$$

$$r_t = \phi_{0,n} + \phi_{1,n}r_{t-1} + \dots + \phi_{n,n}r_{t-n} + \varepsilon_{n,t} \quad (n)$$

Podemos estimar os parâmetros usando o método dos MQ. A idéia da f.a.c.p. é semelhante a idéia de se usar o teste F na regressão múltipla. A estimativa $\hat{\phi}_{1,1}$ é o coeficiente de a.c.p. amostral de *lag* 1 de r_t . A estimativa $\hat{\phi}_{2,2}$ é o coeficiente de autocorrelação amostral de *lag* 2 de r_t , etc.

O coeficiente de autocorrelação parcial $\phi_{k,k}$ é o k -ésimo coeficiente em um processo $AR(k)$, isto é:

$$r_t = \phi_{1,k}r_{t-1} + \phi_{2,k}r_{t-2} + \dots + \phi_{k,k}r_{t-k} \quad (4.13)$$

e mede a correlação adicional entre r_t e r_{t-k} depois de feitos os “ajustamentos” devido aos *lags* intermediários.

No modelo (1), se $\hat{\phi}_{1,1}$ é significativo, r_{t-1} é relevante para explicar r_t . O $\hat{\phi}_{2,2}$ representa a contribuição de r_{t-2} além do modelo $AR(1)$, $\hat{\phi}_{3,3}$ é a contribuição de r_{t-3} além do $AR(2)$, etc. Assim, se a série for $AR(p)$, deveremos ter $\hat{\phi}_{p,p}$ significativamente diferente de zero, e os outros $\hat{\phi}_{p+1,p+1}, \dots$, etc, perto de zero. O raciocínio que levou à definição da f.a.c.p. é o seguinte: Muito da correlação entre r_t e r_{t-k} é devida à correlação deste par com as intermediárias, $r_{t-1}, r_{t-2}, \dots, r_{t-k+1}$. A f.a.c.p. faz a “correção” desta relação.

Temos os seguintes resultados: (i) $\hat{\phi}_{p,p} \xrightarrow{t \rightarrow \infty} \phi_p$; (ii) $\hat{\phi}_{l,l} \xrightarrow{t \rightarrow \infty} zero \quad \forall l > p$; (iii) a variância assintótica de $\hat{\phi}_{l,l}$ é $\frac{1}{T}$ para $l > p$, (iv) Para T grande, $\hat{\phi}_{j,j} \sim \text{Normal}(0, \text{var}(\hat{\phi}_{j,j}))$. Assim, $\hat{\phi}_{j,j}$ será significativamente diferente de zero se $|\hat{\phi}_{j,j}| > \frac{2}{\sqrt{T}}$, $j > p$. Em resumo, a f.a.c.p. “zera” para $l > p$ e é útil.

Suponha uma série de retornos com $T = 864$. Então os limites do intervalo de confiança a 95% são ± 0.06 . Supondo que os valores estimados para os $\phi_{j,j}$ sejam aqueles dados na Tabela 4.1. Escolha uma ordem adequada p para ajuste de $AR(p)$.

Tabela 4.1: Estimativas dos coeficientes de autocorrelação parcial, $\hat{\phi}_{j,j}$, para $j = 1, \dots, 10$.

p	1	2	3	4	5	6	7	8	9	10
f.a.c.p.	0.09	-0.01	-0.11	0.03	0.07	-0.05	0.02	0.06	0.07	-0.02

Para a estimação de um modelo $AR(p)$, pode-se usar condicionalmente as p observações, o método de MQ, isto é, o modelo (4.11), para $t = p + 1, \dots, T$ é visto como um modelo de regressão. O modelo ajustado é

$$\hat{r}_t = \hat{\phi}_0 + \hat{\phi}_1 r_{t-1} + \dots + \hat{\phi}_p r_{t-p}$$

e os resíduos são

$$\hat{a}_t = r_t - \hat{r}_t$$

Da série de resíduos $\{\hat{a}_t\}$ obtemos uma estimativa para σ_a^2 :

$$\hat{\sigma}_a^2 = \frac{\sum_{t=p+1}^T \hat{a}_t^2}{T - 2p - 1},$$

onde o denominador é $T - 2p - 1$ porque temos $p + 1$ parâmetros a estimar e perdemos p observações. O pacote deve dar os erros padrões de $\hat{\phi}_0, \hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\sigma}_a^2$ e devemos fazer o teste t para testar a significância dos parâmetros. Notemos que os $\hat{\phi}_i$'s em finanças são em geral “pequenos”, mas estatisticamente significativos. Alguns pacotes permitem ajustar um modelo $AR(p)$ com alguns parâmetros intermediários ϕ_j iguais a zero.

Alternativamente, pode-se obter estimadores do modelo $AR(p)$ a partir das equações de Yule-Walker, dadas em função das autocovariâncias

$$\gamma_0 = \phi_0 \gamma_1 + \phi_2 \gamma_2 + \dots + \phi_p \gamma_p + \sigma_a^2$$

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \dots + \phi_p \gamma_{k-p}$$

Para estimação dos parâmetros ϕ_j utilizamos as autocovariâncias amostrais (Brockwell, P. J. e Davis, R. A. (1991)).

Para a escolha do melhor modelo, podemos usar algum critério. Todos os critérios de informação existentes na literatura tem como base a verossimilhança. O critério de Akaike, 1973, é definido como

$$AIC = -\frac{2}{T} \mathcal{LL} + \frac{2}{T} k$$

onde \mathcal{LL} representa a log-verossimilhança avaliada nas estimativas de máxima verossimilhança e k representa o número de parâmetros do modelo. No caso de um modelo $AR(p)$ Gaussiano, o AIC reduz a

$$AIC = \ln(\hat{\sigma}_a^2) + 2\frac{p}{T}$$

onde $\hat{\sigma}_a^2$ é a estimativa de máxima verossimilhança de σ_a^2 .

Para a escolha final do melhor modelo verifique a significância das estimativas dos parâmetros, use algum critério e parcimônia. Além disto, após a aceitação do modelo devemos passar à verificação do ajuste do modelo, a qual deve ser sempre cuidadosa e ter por base as suposições feitas.

Por exemplo, devemos observar os gráficos da f.a.c. e f.a.c.p. da série de resíduos $\{\hat{a}_t\}$, que não deve ter coeficientes significativos já que eles são um RB. O teste de Ljung-Box deve ser aplicado na série $\{\hat{a}_t\}$, e procuramos por um p-valor alto, para não rejeitar $H_0: \rho_1 = \dots = \rho_m = 0$. Notemos

que se o modelo é $AR(p)$, então $Q^*(m) \sim \chi_{m-p}^2$ pois p parâmetros foram estimados. Em geral usa-se $m \approx \ln(T)$.

Para previsões de um modelo $AR(p)$, suponha que o tempo corrente é h e o interesse é prever para $h+k$, $k \geq 1$. Então, h é o tempo origem das previsões e k é o horizonte de previsão. Denotaremos por $\hat{r}_h(k)$ a previsão de r_{h+k} usando a função de perda quadrática, isto é, $\hat{r}_h(k)$ satisfaz

$$E[(r_{h+k} - \hat{r}_h(k))^2] \leq \min_g E[(r_{h+k} - g)^2].$$

Notemos que g é uma função que usa a informação disponível até (inclusive) o tempo h . Chamamos o $\hat{r}_h(k)$ a previsão k -passos-a-frente de r_t na origem h .

Previsão 1-passo-a-frente:

Do modelo $AR(p)$ obtemos

$$r_{h+1} = \phi_0 + \phi_1 r_h + \phi_2 r_{h-1} + \dots + \phi_p r_{h-p+1} + a_{h+1}.$$

Se a função de perda for a quadrática, dado o modelo acima, e condicionalmente à toda a informação que temos até o tempo h , temos que $\hat{r}_h(k)$ deve ser a esperança condicional

$$\hat{r}_h(1) = E[r_{h+1} | r_h, r_{h-1}, \dots]$$

$$\hat{r}_h(1) = \phi_0 + \sum_{i=1}^p \phi_i r_{h+1-i}$$

e o erro de previsão associado é $e_h(1) = r_{h+1} - \hat{r}_{h+1} = a_{h+1}$. Temos então que a variância do erro de previsão 1-passo-a-frente é $var(e_h(1)) = var(a_{h+1}) = \sigma_a^2$.

Se, por exemplo, $a_t \sim \text{Normal}$, então o IC 95% para r_{h+1} será $[\hat{r}_h(1) \pm 1.96 \sigma_a]$. Este intervalo acima é considerado um IC condicional, dadas as estimativas, pois não leva em conta a variabilidade das estimativas pontuais $\hat{r}_h(1)$.

Previsão 2-passos-a-frente:

Do modelo $AR(p)$ obtemos

$$r_{h+2} = \phi_0 + \phi_1 r_{h+1} + \phi_2 r_h + \dots + \phi_p r_{h+2-p} + a_{h+2}.$$

A esperança condicional é

$$\hat{r}_h(2) = E[r_{h+2} | r_h, r_{h-1}, \dots] = \phi_0 + \phi_1 \hat{r}_h(1) + \phi_2 r_h + \dots + \phi_p r_{h+2-p}$$

e o erro de previsão associado é

$$e_h(2) = r_{h+2} - \hat{r}_h(2) = \phi_1(r_{h+1} - \hat{r}_h(1)) + a_{h+2}$$

$$e_h(2) = a_{h+2} + \phi_1 a_{h+1}.$$

A variância do erro de previsão é $\text{var}(e_h(2)) = (1 + \phi_1^2)\sigma_a^2$. O IC condicional é calculado da mesma forma, $\hat{r}_h(2) \pm 1.96\sigma_a\sqrt{1 + \phi_1^2}$, dadas as estimativas de máxima verossimilhança. Interessante notar que $\text{var}(e_h(2)) \geq \text{var}(e_h(1))$ significando que a incerteza aumenta.

Previsão vários passos-a-frente

Em geral temos:

$$r_{h+l} = \phi_0 + \phi_1 r_{h+l-1} + \dots + \phi_p r_{h+l-p} + a_{h+l}$$

novamente, a previsão l -passos-a-frente, baseada na função de perda quadrática, é a esperança condicional (dados $r_h, r_{h-1}, r_{h-2}, \dots$) e é dada por

$$\hat{r}_h(l) = \phi_0 + \sum_{i=1}^p \phi_i \hat{r}_h(l-i)$$

onde $\hat{r}_h(j) = r_{h+j}$, se $j \leq 0$. Esta previsão é calculada recursivamente usando as previsões $\hat{r}_h(j)$ para $j = 1, \dots, l$. O erro de previsão l -passos-a-frente é $e_h(l) = r_h(l) - \hat{r}_h(l)$. Pode-se mostrar que para um modelo AR(p) estacionário, temos

$$\hat{r}_h(l) \xrightarrow[l \rightarrow \infty]{} E(r_t),$$

reversão para a média não condicional do processo. Isto é, a previsão a longo prazo converge para a média do processo, e também a $\text{var}(e_h(l)) \xrightarrow[l \rightarrow \infty]{} \text{var}(r_t)$.

Usamos a série do exemplo anterior, com $T = 864$ observações e ajustamos um AR(5). A média μ foi estimada como 0.0098 e $\hat{\sigma}_a = 0.055$. As estimativas dos coeficientes $\hat{\phi}_j$ eram pequenas, e portanto a reversão para a média foi rápida como mostra a Tabela 4.2.

Tabela 4.2: Previsões 1, ..., 6-passos-a-frente a partir do ajuste do modelo AR(5).

passo	1	2	3	4	5	6
previsão	0.0071	-0.0008	0.0086	0.0154	0.0141	0.01
erro padrão (desvio padrão)	0.0541	0.0545	0.0545	0.0549	0.0549	0.0550

4.4.2 Processos Médias Móveis MA(q)

Temos duas abordagens para definir os modelos Médias Móveis de ordem q , MA(q). Os modelos MA podem ser vistos como caso particular do modelo linear (filtro linear) baseado na sequência de choques $\{a_t\}$, $a_t \sim \text{RB}(0, \sigma_a^2)$, ou como um AR(∞) com algumas restrições. Veremos os MA como AR(∞), que favorece o processo de estimação. Um modelo AR(∞) é

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \dots + a_t.$$

Claro que não é realístico pois temos um número infinito de parâmetros para estimar. Para resolver isto propomos restrições, e *uma* das idéias é propor:

$$r_t = \phi_0 - \theta_1 r_{t-1} - \theta_1^2 r_{t-2} - \theta_1^3 r_{t-3} - \cdots + a_t, \quad (4.14)$$

onde temos então apenas *um único* parâmetro a estimar, o θ_1 , pois os coeficientes ϕ_i são:

$$\phi_i = -\theta_1^i, \quad i \geq 1.$$

Para (4.14) ser estacionário devemos ter $|\theta_1| < 1$, porque caso contrário, tanto o θ_1^i quanto a série teriam comportamento explosivo. Sendo $|\theta_1| < 1$, temos que $\theta_1^i \rightarrow 0$, quando $i \rightarrow \infty$, e então a contribuição de r_{t-i} para o r_t decai exponencialmente quando $i \uparrow \infty$, o que é razoável.

O modelo (4.14) pode ser escrito em um formato mais compacto. Para ver isto, reescrevemos o modelo no tempo t

$$r_t + \theta_1 r_{t-1} + \theta_1^2 r_{t-2} + \cdots = \phi_0 + a_t, \quad (4.15)$$

e reescrevemos (4.15) no tempo $t-1$

$$r_{t-1} + \theta_1 r_{t-2} + \theta_1^2 r_{t-3} + \cdots = \phi_0 + a_{t-1}. \quad (4.16)$$

Multiplicando (4.16) por θ_1 e subtraindo o resultado de (4.15) obtemos

$$r_t = \phi_0(1 - \theta_1) + a_t - \theta_1 a_{t-1},$$

o que nos mostra que, a menos da constante $\phi_0(1 - \theta_1)$, a série r_t é uma soma ponderada dos choques nos tempos t e $t-1$, e, por isto é chamado de MA(1), média móvel de ordem 1.

O modelo MA(1) é geralmente escrito como

$$r_t = c_0 + a_t - \theta_1 a_{t-1}, \quad (4.17)$$

onde c_0 é uma constante e $a_t \sim \text{RB}(0, \sigma_a^2)$.

Temos $E[r_t] = c_0$, e $\text{var}(r_t) = \sigma_a^2(1 + \theta_1^2)$. Para ver que os processos MA(1) são inversíveis, considere

$$r_t = (1 - \theta_1 B)a_t. \quad (4.18)$$

Invertendo (4.18) obtemos

$$a_t = (1 - \theta_1 B)^{-1} r_t.$$

Sabemos que se $(1 - \theta_1 B)$ for inversível, pode ser escrito como um polinômio infinito, e igual a $(1 + \theta_1 B + \theta_1^2 B^2 + \theta_1^3 B^3 + \cdots)$. A partir deste fato e sabendo que para o AR(∞) devemos ter $|\theta_1| < 1$, costuma-se chamar esta condição ($|\theta_1| < 1$) de condição de invertibilidade do MA(1).

Para calcular a função de autocorrelação de um MA(1), assumamos, por simplicidade, que $c_0 = 0$. Queremos calcular a $\text{cov}(r_t, r_{t-k})$, $k \geq 1$. Multipliquemos (4.17) por r_{t-k} , e tomemos as esperanças. Temos

$$E[r_t \cdot r_{t-k}] = E[a_t \cdot r_{t-k} - \theta_1 a_{t-1} r_{t-k}]$$

Já que $E[r_t] = 0 \forall t$. Assim, $cov(r_t, r_{t-k}) = cov(r_t, r_{t+k}) = E[r_t \cdot r_{t-k}]$.

$$\gamma_k = E[a_t \cdot r_{t-k}] - \theta_1 E[a_{t-1} r_{t-k}] .$$

Notemos que r_{t-k} é função de a_{t-k} e de a_{t-k-1} e que $E[a_t \cdot r_{t-k}] = 0$ em todos os casos considerando $k \geq 1$. Então $\gamma_k = -\theta_1 E[a_{t-1} r_{t-k}]$. Quando $k = 1$ temos que $a_{t-1} r_{t-k} = a_{t-1}(a_{t-1} - \theta_1 a_{t-2})$ e ficamos com $E[(a_{t-1})^2]$. Então $\gamma_1 = -\theta_1 \sigma_a^2$. Se $k \geq 2$, a esperança $E[a_{t-1} r_{t-k}]$ é zero. Resumindo, para um MA(1),

$$\gamma_k = -\theta_1 \sigma_a^2 \quad \text{se } k = 1$$

$$\gamma_k = 0 \quad \text{se } k \geq 2.$$

Como $\gamma_0 = var(r_t) = \sigma_a^2(1 + \theta_1^2)$, obtemos, para o MA(1)

$$\rho_0 = 1, \quad \rho_1 = \frac{-\theta_1}{1 + \theta_1^2}, \quad e \quad \rho_k = 0, \quad k \geq 2 \quad (4.19)$$

e vemos então que a f.a.c. de um MA(1) é muito útil para identificar um modelo (tem a forma de uma f.a.c.p. de um AR(1)). A memória deste processo é de apenas 1 período, e sua f.a.c. “zera” a partir do lag 2.

Notemos que $\rho_1 = \frac{-\theta_1}{1 + \theta_1^2} \Rightarrow \rho_1 + \rho_1 \theta_1^2 + \theta = 0$, uma função quadrática em θ_1 , para a qual $\Delta = 1 - 4\rho_1^2 > 0$, pois θ_1 tem que ser real. Isto implica em $\rho_1^2 < 1/4 \Rightarrow -\frac{1}{2} < \rho_1 < \frac{1}{2}$. Considere novamente a quadrática e a reparametrização $1/\theta$. Temos, $\rho_1 \frac{1}{\theta_1^2} + \frac{1}{\rho_1} + \rho_1 = 0$, e portanto $\rho_1 \theta_1^2 + \theta + \rho_1 + \rho_1 = 0$. Isto é, dado ρ_1 , dois valores θ_1 e $\frac{1}{\theta_1}$ satisfazem a equação e portanto temos dois processos MA(1) com a mesma f.a.c.

O modelo MA(2) tem a forma

$$r_t = c_0 + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}. \quad (4.20)$$

Analogamente, pode-se mostrar que para um MA(2)

$$\rho_0 = 1, \quad \rho_1 = \frac{-\theta_1 + \theta_1 \theta_2}{1 + \theta_1^2 + \theta_2^2}, \quad \rho_2 = \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2}, \quad e \quad \rho_k = 0 \quad \text{para } k \geq 3 \quad (4.21)$$

então a memória de um MA(2) é finita. A Figura 4.7 mostra as f.a.c.'s de um MA(1) com $\theta_1 = 0.25$ e de um MA(2) com $\theta_1 = 0.25$ e $\theta_2 = -0.25$.

Um modelo MA(q) tem a forma

$$r_t = c_0 + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad q > 0, \quad (4.22)$$

onde c_0 é uma constante e $a_t \sim \text{RB}(0, \sigma_a^2)$. Para as restrições no espaço paramétrico dos modelos MA(q) veja Hamilton, J. D. (1994).

A forma geral de um MA(q) usando operadores retroativos é:

$$r_t = \Theta(B)a_t$$

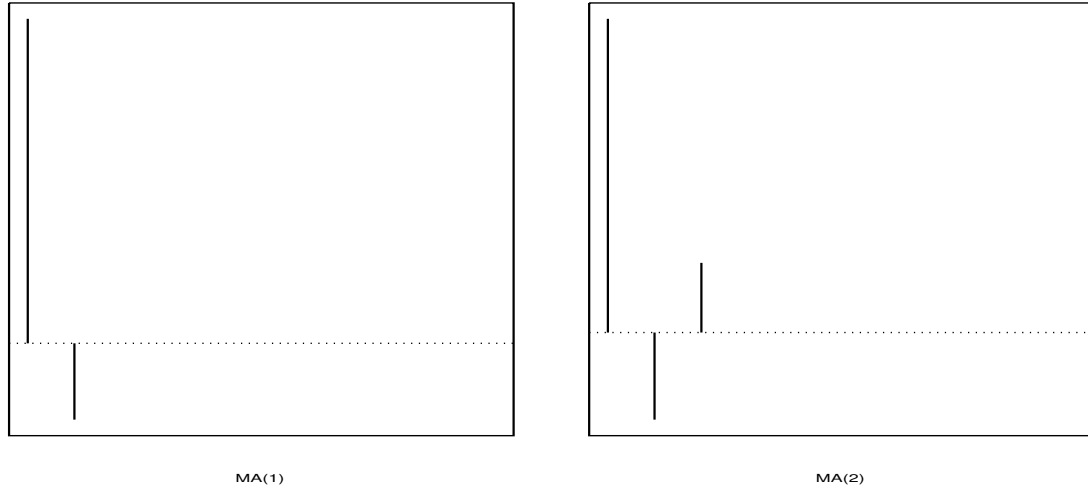


Figura 4.7: *f.a.c.s* de um $MA(1)$ com $\theta_1 = 0.25$ e de uma $MA(2)$ com $\theta_1 = 0.25$ e $\theta_2 = -0.25$.

onde o polinômio $\Theta(B)$ de ordem q é

$$\Theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 - \dots - \theta_q B^q).$$

Para quaisquer valores $\theta_1, \theta_2, \dots, \theta_q$, os processos $MA(q)$ são sempre fracamente estacionários porque são combinações lineares de um número finito de choques que são $RB(0, \sigma_a^2)$, e portanto possuem os dois primeiros momentos invariantes no tempo. Um processo $MA(q)$ tem esperança e variância não-condicionais dadas por

$$\mu \equiv E[r_t] = c_0$$

$$\gamma_0 \equiv \text{var}(r_t) = E[(r_t - \mu)^2] = (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) \sigma_a^2$$

e covariâncias dadas por

$$\gamma_k \equiv \text{cov}(r_t, r_{t-k}) = 0$$

para $k > q$; e para $k = 1, \dots, q$,

$$\gamma_k = (\theta_k + \theta_{k+1}\theta_1 + \theta_{k+2}\theta_2 + \dots + \theta_q\theta_{q-k}) \sigma_a^2.$$

Como $\sum_{j=-\infty}^{\infty} |\gamma_j| < \infty$, temos ergodicidade para a média. Se processo Gaussiano teremos ergodicidade para todos os momentos.

Para a identificação das ordens dos modelos médias móveis, a f.a.c. é muito útil. Observe, por exemplo, a Figura 4.8. Nesta figura vemos as f.a.c.'s amostrais dos índices dos mercados da Indonésia e do México, as quais parecem indicar, respectivamente, um MA(1) e um MA(2). Assim, poderíamos propor o modelo $r_t = c_0 + a_t - \theta_1 a_{t-1}$ para a Indonésia, e o modelo $r_t = c_0 + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$ para o México.

Para a estimação, o método mais utilizado é o da máxima verossimilhança. Temos duas abordagens, condicional e exata. Considere um MA(q), $r_t = c_0 + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$. A abordagem condicional assume que os choques iniciais são zero, isto é, $a_0 = 0 = a_{-1} = a_{-2} \dots$. Assim, $r_1 = c_0 + a_1$, e portanto $a_1 = r_1 - c_0$. A seguir temos $r_2 = c_0 + a_2 - \theta_1 a_1$. A função de verossimilhança é então escrita condicionalmente para as v.a.'s a_t , sabendo que $a_1 = r_1 - c_0$, $a_2 = r_2 - c_0 + \theta_1 a_1$, etc. A maximização desta função dá origem aos EMV condicionais.

Os EMV exatos tratam os choques iniciais a_t , $t \leq 0$, como parâmetros extras desconhecidos, e os estimam. É um método computacionalmente mais complexo. Se T é grande, os 2 métodos resultam em estimativas parecidas.

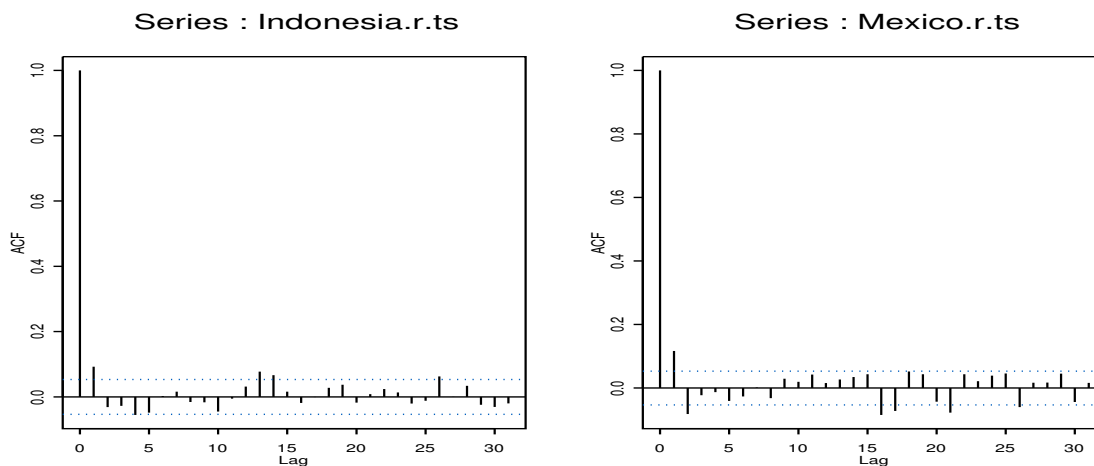


Figura 4.8: f.a.c.'s amostrais das séries de retornos do índice da Indonésia e do México.

Exemplo: México. Iremos inicialmente estimar o modelo mais completo, o modelo $r_t = c_0 + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$.

```
ARIMA order:  0 0 2
      Value Std. Error t-value
ma(1) -0.12620    0.02706  -4.664
ma(2)  0.07888    0.02706   2.915
Variance-Covariance Matrix:
```

```

          ma(1)      ma(2)
ma(1) 0.0007323342 0.0001003436
ma(2) 0.0001003436 0.0007323342
Estimated innovations variance: 4.4755
Optimizer has converged
AIC: 5888.6517

```

```

Ajuste MA(1)
ARIMA order: 0 0 1
      Value Std. Error t-value
ma(1) -0.1411    0.02687   -5.25

```

```

Variance-Covariance Matrix:
          ma(1)
ma(1) 0.0007222483
Estimated innovations variance: 4.5027
Optimizer has converged
AIC: 5894.8788 (maior!)

```

A Figura 4.9 mostra os log-retornos diários do México e os resíduos do ajuste MA(2) aos mesmos.

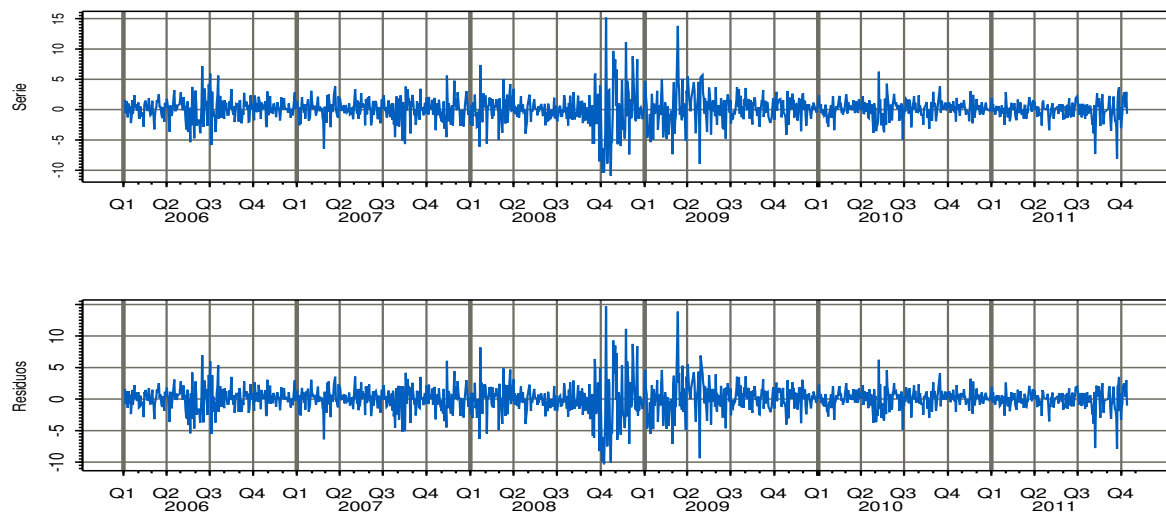


Figura 4.9: *Série de retornos e resíduos do ajuste MA(2) aos log-retornos do Mexico.*

Como a memória é finita, as previsões feitas a partir de um $MA(q)$ reverterem para a média $E[r_t]$ rapidamente.

Previsão 1-passo-a-frente de um $MA(1)$: Seja h a origem das previsões. O retorno no período seguinte é dado por

$$r_{h+1} = c_0 + a_{h+1} - \theta_1 a_h \quad (4.23)$$

Tomando a esperança condicional (dado todo o passado até a origem h), temos que

$$\hat{r}_h(1) = E[r_{h+1} \mid r_h, r_{h-1}, \dots] = c_0 - \theta_1 a_h$$

e sendo

$$e_h(1) = r_{h+1} - \hat{r}_h(1) = a_{h+1}$$

temos que a variância do erro de previsão um passo a frente é

$$\text{var}(e_h(1)) = \sigma_a^2.$$

Na prática, para obter o a_h na (4.23) é comum supor $a_0 = 0$, calcular $\hat{a}_1 = r_1 - \hat{c}_0$, e continuar calculando \hat{a}_t , para $2 \leq t \leq h$, recursivamente usando

$$a_t = r_t - c_0 + \theta_1 a_{t-1}.$$

Os parâmetros c_0 e θ_1 são substituídos por suas estimativas \hat{c}_0 e $\hat{\theta}_1$.

Previsão 2-passos-a-frente de um $MA(1)$: Seja h a origem das previsões. Em $h + 2$ temos:

$$r_{h+2} = c_0 + a_{h+2} - \theta_1 a_{h+1} \quad (4.24)$$

e então a previsão é

$$\hat{r}_h(2) = E[r_{h+2} \mid r_h, r_{h-1}, \dots] = c_0$$

e o erro de previsão é

$$e_h(2) = r_{h+2} - \hat{r}_h(2) = a_{h+2} - \theta_1 a_{h+1}$$

e portanto a variabilidade do erro de previsão é dada por

$$\text{var}(e_h(2)) = \sigma_a^2(1 + \theta_1^2).$$

Compare com a variância do erro de previsão 1-passo-a-frente, que é σ_a^2 . A variância do erro no passo $h + 2$ é sempre maior ou igual a $\text{var}(e_h(1))$. Notemos que a previsão 2-passos-a-frente de um $MA(1)$ é simplesmente a média (não condicional) do processo, c_0 . Assim, para um $MA(1)$, as previsões para horizontes mais longos são dadas por

$$\hat{r}_h(l) = c_0, \quad \forall l \geq 2.$$

Previsões para um $MA(2)$: Seja h a origem das previsões, e considere o período $h + 1$.

$$r_{h+l} = c_0 + a_{h+l} - \theta_1 a_{h+l-1} - \theta_2 a_{h+l-2}. \quad (4.25)$$

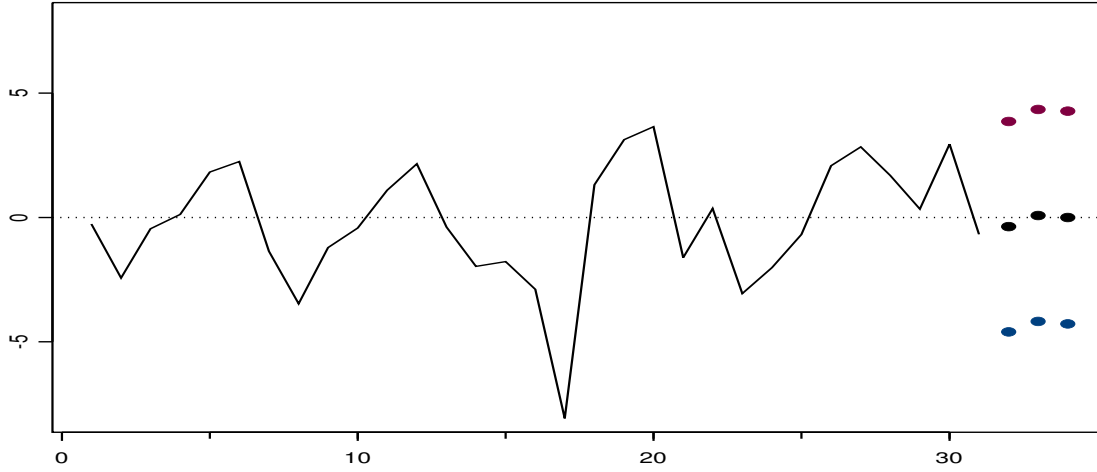


Figura 4.10: Previsões baseadas no ajuste $MA(2)$ para os retornos do Mexico.

As previsões são dadas por $\hat{r}_h(1) = c_0 - \theta_1 a_h - \theta_2 a_{h-1}$, $\hat{r}_h(2) = c_0 - \theta_2 a_h$, $\hat{r}_h(l) = c_0$ para $l \geq 3$, e vemos que novamente as previsões reverterem para a média não condicional do processo, c_0 .

Os erros de previsão são $e_h(2) = r_{h+2} - \hat{r}_h(2) = a_{h+2} - \theta_1 a_{h+1}$, e portanto a variabilidade do erro de previsão é dada por $var(e_h(2)) = \sigma_a^2(1 + \theta_1^2)$. Temos que a $var(e_h(l)) \rightarrow \gamma_0$, γ_0 a variância não condicional da série, após 2 passos ($l > 2$).

Pode-se mostrar que para um $MA(q)$, a previsão l -passos-a-frente será c_0 para $l > q$, e a variância do erro se iguala a variância da série após q passos. Resumindo, tanto para o AR quanto para o MA, as previsões irão eventualmente convergir para a média não condicional, e a variância do erro de previsão irá convergir para a variância (não condicional) da série.

Exemplo: Mexico-continuação. Previsões 1, 2, e 3 passos a frente para o modelo $MA(2)$. Os erros padrões das previsões são, respectivamente, 2.115, 2.132 e 2.138, e mostramos o IC correspondente na Figura 4.10.

4.4.3 Processos ARMA(p, q)

Considere o modelo $ARMA(p, q)$

$$r_t = \phi_0 + \sum_{i=1}^p \phi_i r_{t-i} + a_t - \sum_{i=1}^q \theta_i a_{t-i} \quad (4.26)$$

onde $\{a_t\} \sim RB(0, \sigma_a^2)$, p e q inteiros não negativos. Já vimos os casos especiais $AR(p) = ARMA(p, 0)$ e $MA(q) = ARMA(0, q)$.

Usando o operador retroativo B podemos escrever

$$(1 - \phi_1 B - \dots - \phi_p B^p)r_t = \phi_0 + (1 - \theta_1 B - \dots - \theta_q B^q)a_t$$

ou

$$\Phi(B)r_t = \phi_0 + \Theta(B)a_t.$$

Do lado esquerdo temos o polinômio do modelo $\text{AR}(p)$, e do lado direito o polinômio do modelo $\text{MA}(q)$. A caracterização do $\text{ARMA}(p, q)$ como um $\text{AR}(p)$ introduz a equação característica do modelo ARMA . Se todas as raízes desta equação forem em módulo menores que um, então o processo é fracamente estacionário. Em um ARMA geral, a condição suficiente para inversibilidade é que as raízes do polinômio $\Theta(B) > 1$ em módulo. Isto é equivalente a ter $|\theta| < 1$. A média não condicional do processo é $E[r_t] = \frac{\phi_0}{1 - \phi_1 - \dots - \phi_p} \equiv \mu$. A esperança condicional depende dos p retornos passados e a variância condicional é σ_a^2 .

Considere os polinômios

$$\Phi(B) = 1 - \sum_{i=1}^p \phi_i B^i$$

$$\Theta(B) = 1 - \sum_{i=1}^q \theta_i B^i$$

Sabemos que um polinômio infinito pode ser escrito como a razão entre dois polinômios finitos (e vice-versa) e portanto

$$\frac{\Theta(B)}{\Phi(B)} = 1 + \psi_1 B + \psi_2 B^2 + \dots \equiv \Psi(B) \quad (4.27)$$

e

$$\frac{\Phi(B)}{\Theta(B)} = 1 - \pi_1 B - \pi_2 B^2 - \dots \equiv \Pi(B) \quad (4.28)$$

Notemos que, por definição $\Psi(B)\Pi(B) = 1$, e que $\frac{\phi_0}{\Theta(1)} = \frac{\phi_0}{1 - \theta_1 - \dots - \theta_q}$ e que $\frac{\phi_0}{\Phi(1)} = \frac{\phi_0}{1 - \phi_1 - \dots - \phi_p}$.

(i) Representação AR . Lembrando que $(1 - \phi_1 B - \dots - \phi_p B^p)r_t = \phi_0 + (1 - \theta_1 B - \dots - \theta_q B^q)a_t$ (ou, $\Theta(B)r_t = \phi_0 + \Phi(B)a_t$), e a partir de (4.28) temos a representação $\text{AR}(\infty)$ do $\text{ARMA}(p, q)$:

$$\frac{\Phi(B)}{\Theta(B)}r_t = \frac{\phi_0}{\Theta(B)} + a_t$$

$$(1 - \pi_1 B - \pi_2 B^2 - \dots)r_t = \frac{\phi_0}{\Theta(B)} + a_t$$

ou

$$r_t = \frac{\phi_0}{\Theta(B)} + \pi_1 r_{t-1} + \pi_2 r_{t-2} + \dots + a_t$$

$$r_t = \frac{\phi_0}{1 - \theta_1 - \dots - \theta_q} + \pi_1 r_{t-1} + \pi_2 r_{t-2} + \dots + a_t,$$

representação que enfatiza dependência do log-retorno corrente nos outros passados. Essa dependência é dada pelos pesos π . Os π_i devem decair para zero quando i cresce, para que a contribuição desapareça

no tempo. Neste caso, isto é, quando $\pi_i \rightarrow 0$ quando $i \uparrow$, o ARMA é dito inversível. Exemplo: Um AR puro tem $\Theta(B) = 1$, e portanto $\Pi(B) = \Phi(B)$, e os pesos π são em número finito, e o modelo é inversível.

Resumindo, para um ARMA(p, q), $\Phi(B)r_t = \Theta(B)a_t$, temos que a condição de estacionariedade é dada pela condição de estacionariedade de um AR(p); condição de inversibilidade é dada pela condição de inversibilidade de um MA(q). Sua f.a.c. parece com a f.a.c. de um AR(p). Sua f.a.c.p. se parece com a f.a.c.p. de um MA(q). Ambas são de difícil interpretação. Tsay, R. S. e Tiao, G. C. (1984) propõem a função f.a.c.e., função de autocorrelação estendida (ver também Tsay, R. S. (2005)). Para a escolha das ordens do modelo o melhor é ajustar vários modelos simples e examinar os valores do AIC.

O modelo utilizado mais frequentemente para modelar retornos financeiros é o ARMA(1, 1):

$$r_t = \phi r_{t-1} + a_t - \theta a_{t-1} \quad (4.29)$$

Para este modelo,

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{(1 - \phi\theta)(\phi - \theta)}{1 + \theta^2 - 2\phi\theta}$$

e

$$\rho_k = \phi \rho_{k-1} \quad \text{para } k > 1.$$

Em geral, retornos financeiros apresentam ambos ϕ e θ positivos com $\phi > \theta$. Se $(\phi - \theta)$ for pequeno, o ρ_1 pode ser bem menor que ϕ , mas o AR(1) tem $\rho_1 = \phi$.

O critério AIC e BIC para a identificação do melhor modelo consiste na escolha das ordens k e l que minimizam a quantidade

$$P(k, l) = -\ln \hat{\sigma}_{k,l}^2 + (k + l) \frac{C(T)}{T}$$

onde $\hat{\sigma}_{k,l}^2$ é uma estimativa da variância residual obtida ajustando um modelo ARMA(k, l) às T observações da série e $C(T)$ é função do tamanho da amostra T . Note que $(k + l) \frac{C(T)}{T}$ é o termo penalizador, que aumenta com o número de parâmetros. Por outro lado, se o número de parâmetros cresce, a variância dos resíduos diminui. Assim, minimizar $P(k, l)$ é encontrar as ordens k e l que equilibrem este comportamento.

(i) Critério de informação AIC (Akaike, 1973, 1974)

$$\text{AIC}(k, l) = \frac{2}{T} \ln \hat{\sigma}_{k,l}^2 + \frac{2(k + l)}{T}$$

onde $\hat{\sigma}_{k,l}^2$ é o EMV de σ_a^2 para o modelo ARMA(k, l). Na prática, estipulamos um valor máximo K para k e um L para l e olhamos todas as combinações (k, l) possíveis. Alguns textos sugerem usar $K = L = \ln T$.

Existem várias correções para melhorar o critério AIC, no sentido de diminuir a probabilidade de selecionar uma ordem maior que a verdadeira. Hurvich, C. M. e Tsai, C. L. (1989) propuseram a seguinte correção no caso AR(k), adequada no caso de ser T pequeno

$$\text{AIC}_c(k) = \text{AIC}(k) + \frac{2(k + 1)(k + 2)}{T - k + 2}, \quad k \leq K$$

(ii) Critério de informação Bayesiano BIC

$$\text{BIC}(k, l) = -2 \ln(\text{FV}) + (\# \text{ parâmetros}) \ln(T)$$

onde FV é a função de verossimilhança. No caso $\text{ARMA}(k, l)$ fica:

$$\text{BIC}(k, l) = \ln \hat{\sigma}_{k,l}^2 + (k + l) \frac{\ln T}{T}.$$

Exemplo. Usando a série de log-retornos do Mexico, fizemos $K = L = 8$, isto é, ajustamos todas as combinações de modelos para $i, j = 1, 2, \dots, 8$. A tabela abaixo mostra os AIC's desses modelos. O menor AIC indica $p = 8$ e $q = 0$. Contudo, o ajuste deste modelo apresenta várias estimativas de parâmetros não significativas. O modelo foi ser revisto retirando-se um parâmetro de cada vez, e o melhor ajuste foi aquele já obtido, um $\text{AR}(2)$.

```
round(aic, 1)
  q    0      1      2      3      4      5      6      7      8
p 0 ---    5894.9 5888.7 5889.7 5891.6 5892.2 5892.8 5894.2 5894.3
  1 5894.9 5887.6 5885.3 5887.4 5889.5 5889.9 5890.9 5892.4 5891.9
  2 5880.7 5882.7 5883.5 5886.2 5887.6 5888.3 5889.3 5891.0 5891.0
  3 5879.2 5880.8 5882.5 5883.6 5885.4 5886.9 5888.1 5889.7 5889.4
  4 5877.1 5878.2 5880.5 5881.3 5883.5 5885.1 5886.2 5887.7 5887.5
  5 5873.2 5875.2 5877.2 5879.1 5881.0 5882.1 5883.9 5886.0 5885.4
  6 5870.9 5872.9 5874.7 5876.3 5878.7 5880.6 5882.6 5884.2 5882.8
  7 5869.5 5871.5 5872.9 5875.2 5877.2 5879.1 5881.2 5879.2 5879.6
  8 5866.1 5867.6 5869.6 5871.3 5872.7 5874.8 5876.6 5876.7 5878.0
```

```
Call: arima.mle(x = Mexico.r.ts@data, model = list(order = c(8, 0, 0)))
Method: Maximum Likelihood with likelihood conditional on 8 observations
```

```
ARIMA order: 8 0 0
```

	Value	Std. Error	t-value
ar(1)	0.127800	0.02721	4.6960
ar(2)	-0.100000	0.02743	-3.6460
ar(3)	-0.003990	0.02756	-0.1448
ar(4)	-0.015340	0.02754	-0.5571
ar(5)	-0.037990	0.02754	-1.3800
ar(6)	-0.023370	0.02756	-0.8482
ar(7)	0.004419	0.02743	0.1611
ar(8)	-0.038500	0.02721	-1.4150

No caso ARMA é interessante usar os EMV devido às suas boas propriedades assintóticas, embora a suposição da distribuição dos erros tenha que ser feita. Uma dessas propriedades assintóticas é que o vetor de estimadores $(\hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q)$ possui distribuição Normal $_k$, $k = p+q$. Existe covariância diferente de zero entre os elementos de $(\hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q)$, mas este é não correlacionado com $\hat{\sigma}^2$. As previsões baseadas no ARMA(p, q) são aproximadamente as previsões de um AR(p). Para mais detalhes consulte Hamilton, J. D. (1994). Previsão 1-passo-a-frente de um ARMA(p, q) é dada por:

$$\hat{r}_h(1) = E[r_{h+1}|r_h, r_{h-1}, \dots] = \phi_0 + \sum_{i=1}^p \phi_i r_{h+1-i} - \sum_{i=1}^q \theta_i a_{h+1-i}$$

e o erro de previsão correspondente é $e_h(1) = r_{h+1} - \hat{r}_h(1) = a_{h+1}$. A variância do erro da previsão 1-passo-a-frente é $\text{var}(e_h(1)) = \sigma_a^2$. Para as fórmulas da previsão e seu erro para vários passos-a-frente, veja Morettin, P. A. e Toloi, C. M. C. (2006).

4.4.4 Processos ARIMA(p, d, q)

Um processo $\{X_t\}$ segue um modelo ARIMA(p, d, q), d inteiro, se $\Delta^d X_t = r_t$ seguir um ARMA(p, q), isto é,

$$r_t = \phi_0 + \phi_1 r_{t-1} + \dots + \phi_p r_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}, \quad (4.30)$$

onde $\phi_0 = \mu(1 - \phi_1 - \dots - \phi_p)$. Supondo $\mu = 0$, temos $\Phi(B)\Delta^d X_t = \Theta(B)a_t$. Por exemplo, o processo X_t pode ser o processo P_t dos preços de um ativo financeiro (ou o $\log(P_t)$), quando, *em geral*, basta tomar $d = 1$.

Na prática, dada uma série temporal de log-retornos percentuais e após ter sido testada e confirmada a hipótese de estacionariedade, para se estabelecer o melhor modelo ARMA(p, q) para a mesma temos 3 estágios: (1) Identificação do modelo; (2) Estimção do modelo; (3) Diagnósticos. Este ciclo deve ser iterado se no estágio 3 verificarmos que o modelo não é adequado, ou se no estágio 2 ocorrer algum problema de otimização.

1 Identificação. Esta é a fase crítica. A identificação é feita a partir da inspeção das f.a.c. e f.a.c.p. amostrais. Devemos ter em mente as formas das f.a.c.'s e das f.a.c.p.'s teóricas dos modelos estudados. Neste procedimento de identificação devemos

- verificar se há necessidade de transformação da série original, com o objetivo de estabilizar a variância. Por exemplo, em geral tomamos o logaritmo neperiano da série P_t .
- tomar diferenças da série obtida no item anterior, até se obter uma série estacionária. Por exemplo, em geral basta a primeira diferença da série obtida no item anterior. A f.a.c. desta série deve decrescer rapidamente para zero. Deve-se então fazer testes (teste *ADF*) para verificação da existência de raiz unitária. Pode-se também fazer um teste para verificar a existência de d fracionário.
- identificar as ordens p e q do processo *ARMA* analisando as f.a.c. e f.a.c.p. amostrais.

Na prática, no caso de retornos financeiros, d é 0, ou 1, ou no máximo 2, e basta examinar as primeiras 15 ou 20 autocorrelações amostrais. No caso do modelo ARMA, devido à dificuldade de sua

identificação, em geral estimamos alguns modelos simples (ordens p e q entre 0 e 3) e escolhemos o melhor modelo após a estimação usando os critérios AIC ou BIC. Por exemplo, a Figura 4.11 mostra as f.a.c. e f.a.c. parcial amostrais para o índice brasileiro. A figura ilustra bem a dificuldade prática de se decidir sobre o modelo mais adequado. Vemos que nem o decaimento dos $\rho_{j,T}$ nem o decaimento dos $\widehat{\phi}_{j,j}$ indicam claramente qual o melhor modelo. Poderíamos tentar um ARMA(2,2) para esta série.

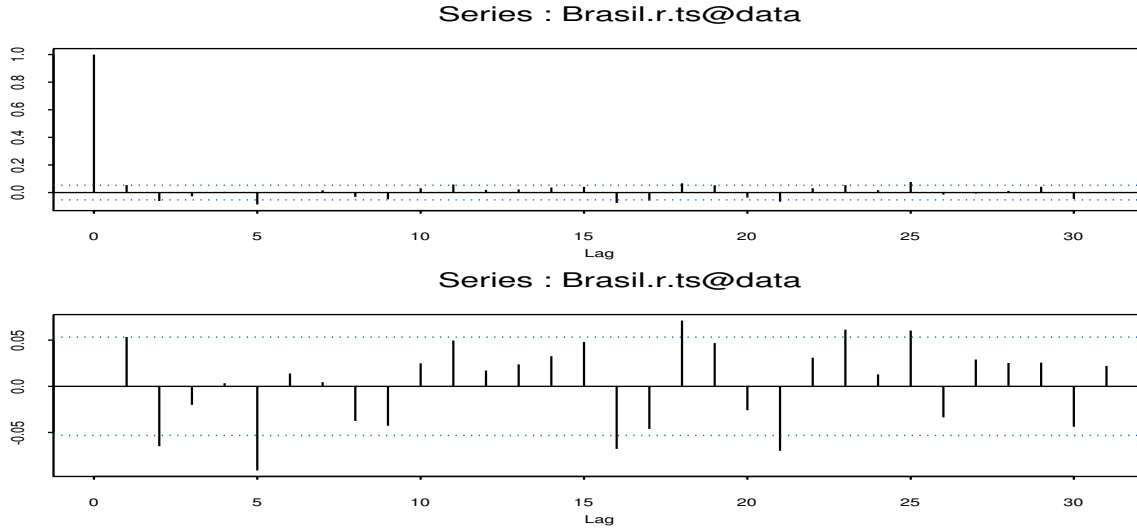


Figura 4.11: Funções de autocorrelação e de autocorrelação parcial amostrais para os log-retornos diários do Brasil.

Por exemplo, para a série da China, obtivemos as estimativas dos coeficientes de autocorrelação serial e parcial

	j	1	2	3	4	5	6	7	8	9
ro-chapeu j		0.06	-0.05	-0.02	-0.04	-0.02	0.01	0.03	0.03	-0.03
fi chapeu jj		-0.06	-0.01	-0.04	-0.01	0.01	0.03	0.03	-0.03	-0.03

Uma estimativa para o erro padrão dos coeficientes $\widehat{\phi}_{jj}$ é $1/\sqrt{T} = 0.0271$, $T = 1357$, e portanto $2\sigma(\widehat{\phi}_{jj}) = 0.0542$. Vemos que nem o decaimento dos $\widehat{\rho}_j$ nem o decaimento dos $\widehat{\phi}_{jj}$ indicam claramente o melhor modelo.

2 Estimação. Pode-se usar o método de Mínimos Quadrados, o método dos Momentos, ou o método da Máxima Verossimilhança condicional ou exata. No caso ARMA é interessante usar os EMV devido às suas boas propriedades assintóticas, embora a suposição da distribuição das inovações a_t tenha que ser feita.

3 *Diagnóstico*. Após estimar, temos que verificar se o modelo representa adequadamente os dados. Qualquer inconsistência descoberta pode sugerir um novo modelo. Para representar adequadamente os dados, o melhor modelo ajustado deve ter todas as estimativas de seus parâmetros significantes (teste t). Todos os parâmetros que não forem estatisticamente significantes devem ser retirados do modelo e um novo ajuste efetuado.

Na verdade, o “super-ajustamento”, isto é, o ajuste de um modelo com vários parâmetros, é uma técnica bastante utilizada. Após o ajuste do super-modelo, verificamos: (i) se os novos parâmetros são significantes; (ii) se a inclusão dos mesmos diminui significativamente a variância dos resíduos.

Suponha que o modelo ajustado tenha sido o ARMA $\Phi(B)r_t = \Theta(B)a_t$, com $r_t = \Delta^d X_t$. Se este for o modelo verdadeiro, e se ele estiver bem estimado, os resíduos $\hat{a}_t = \hat{\Theta}^{-1}(B)\hat{\Phi}(B)r_t$ devem ser RB. Isto é, se o modelo ajustado for o modelo verdadeiro, e se ele estiver bem estimado, os resíduos devem ser ruído branco e portanto espera-se que as estimativas dos coeficientes de autocorrelação estejam próximas de zero para todos os *lags* (dentro do intervalo de confiança). Na prática se usa o intervalo de 95% de confiança, definido pelos valores $\pm 2/\sqrt{T}$, veja detalhes em Durbin, J. (1970). O teste de Ljung-Box deve também ser utilizado. Devemos fazê-lo para o número K de *lags* menor ou igual a $K = 15$ ou 20 , e a distribuição da estatística teste é uma χ^2_{K-p-q} .

No exemplo do ajuste da série do México, apresentamos abaixo o teste de Ljung-Box para 10 *lags*, e para a série de retornos, e depois para os resíduos do modelo AR(2) ajustado.

\$BL.r:

Statistic	P-value	Chi^2-d.f.
34.88955	0.0001304013	10

\$BL.r:

Statistic	P-value	Chi^2-d.f.
7.477963	0.679682	10

O objetivo final quase sempre é o de fazer previsões. Conforme vimos, a previsão obtida pela minimização do erro quadrático médio (EQM) é dada pela esperança condicional. O SPlus e o R fornecem previsões baseadas no critério de EMQ. Devemos nos lembrar que as previsões contém erros.

4.4.5 Modelos Sazonais

Uma série X_t com componente sazonal de *lag* 12 teria sua f.a.c. como no lado direito da Figura 4.12, e um gráfico no tempo com uma tendência suave crescente como no lado esquerdo da mesma figura.

Se tomarmos a primeira diferença desta série poderemos “corrigir” a tendência no tempo, mas isto irá acentuar a significância de autocorrelação de *lag* 12. Veja a Figura 4.13.

Seja a série $Y_t = \Delta X_t$. Para modelar a componente sazonal, aplicamos o operador diferença de ordem 12 obtendo a série Z_t

$$Z_t = \Delta_{12}Y_t = (1 - B^{12})Y_t$$

isto é

$$Z_t = Y_t - Y_{t-12}.$$

A série Z_t terá 12 observações a menos que Y_t , e assim perdemos ao todo 13 observações. A Figura 4.13 mostra a f.a.c. e f.a.c.p. amostrais da série Z_t , onde não vemos mais indicação de efeito sazonal. Nosso modelo para a série Z_t seria então

$$Z_t = (1 - B)(1 - B^{12})X_t$$

Podemos completar o modelo incluindo a parte MA e MA sazonal:

$$(1 - B)(1 - B^S)X_t = (1 - \Theta B)(1 - \Theta_S B^S)a_t,$$

com $|\theta| < 1$ e $|\theta_S| < 1$.

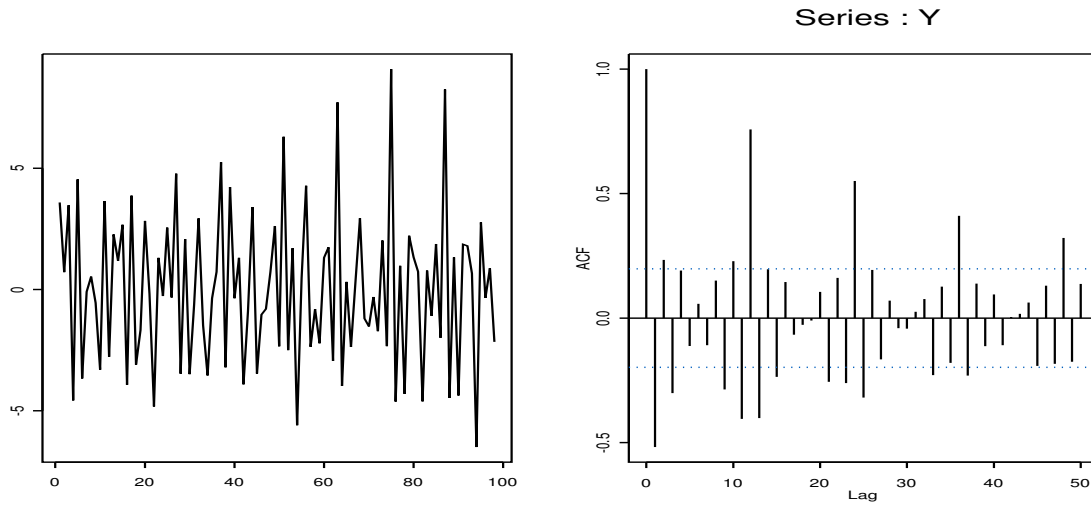


Figura 4.12: *Série diferenciada da série componente sazonal de período 12.*

O modelo acima é um caso particular do SARIMA geral, o ARIMA Sazonal Multiplicativo de ordem $(p, d, q) \times (P, D, Q)$, que pode ser escrito como

$$\phi(B)\Phi(B^S)\Delta^d\Delta_S^D X_t = \theta(B)\Theta(B^S)a_t$$

onde Δ_S é o operador diferença sazonal, D indica quantas diferenças, Φ é o operador autorregressivo estacionário de ordem P , e Θ é o operador média sazonal de ordem Q .

4.4.6 Processos ARFIMA(p, d, q)

A maioria das séries que analisamos são $I(0)$ ou $I(1)$. Por isto vale a pena examinar primeiro com detalhes séries desses dois tipos.

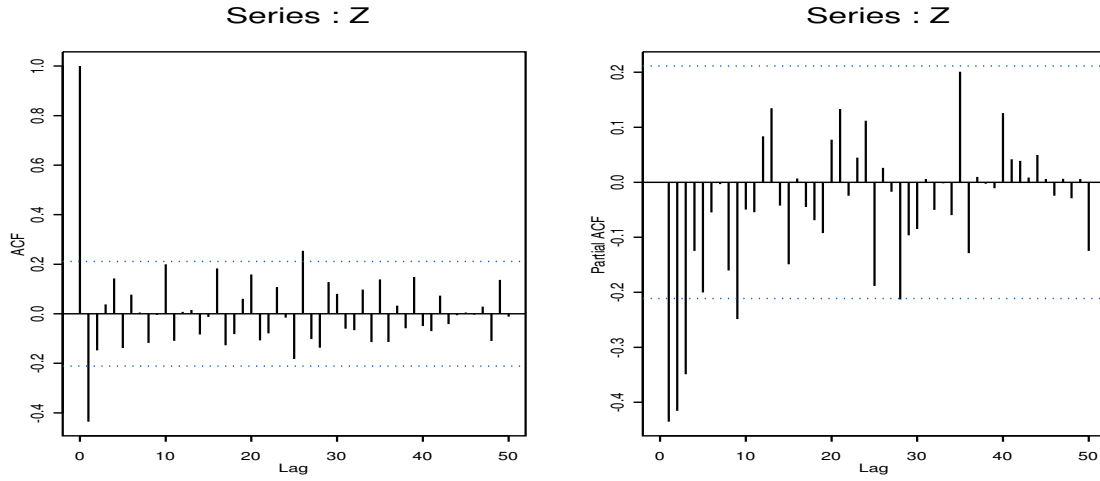


Figura 4.13: Série diferenciada de lag 12 da série anterior.

Se r_t é $I(0)$, denotamos $r_t \sim I(0)$, e assumindo que a média é zero, temos: (i) $\text{var}(r_t) < \infty$ a qual não depende de t ; (ii) as inovações a_t têm efeito temporário em r_t ; (iii) a esperança do tempo entre os cruzamentos do eixo $r = 0$ é finita, ou seja, r_t flutua em torno da média; (iv) as autocorrelações ρ_k decrescem em magnitude quando k cresce, e sua soma é finita.

Mesmo sendo a primeira diferença de uma série estacionária, também estacionária, não devemos super-diferenciar, pois isto levaria a um modelo mais complicado, talvez não inversível, e de grande dificuldade computacional (o processo de estimação). Mais ainda, a variância desta série super-diferenciada será maior. Por isto, é interessante observar o comportamento da $\widehat{\text{var}}(\Delta^d r_t)$ em função de d . Esta estimativa deve decrescer até a série estacionarizar, e depois tende a crescer novamente. Esta é apenas uma ferramenta auxiliar exploratória e não deve ser usada como a única para se determinar o valor certo para d .

Como já observamos, séries de retornos financeiros são em geral $I(0)$ ou $I(1)$. A f.a.c. de uma série $I(1)$ decai lentamente, e pode-se mostrar que para lag k fixo, $\hat{\rho}_k \uparrow 1$ quando T tende a infinito. Por outro lado, como também já ressaltamos, a f.a.c. de uma série $I(0)$ decai exponencialmente para zero a medida que o lag aumenta e assim, observações separadas no tempo por um lag grande podem ser consideradas independentes. Contudo, existem algumas séries para as quais a f.a.c. decai lentamente para zero numa taxa polinomial quando o lag aumenta. Tais processos são ditos de *memória longa*. Memória longa na média condicional de um processo tem sido observada em dados vindo de áreas como meteorologia, astronomia, hidrologia, economia, veja Beran (1994).

Um dos processos apresentando esta característica é o processo auto regressivo-média móvel fracionalmente diferenciado $\text{ARFIMA}(p, d, q)$ que modela a presença de memória longa na média condi-

cional da série (além da memória curta) e pode ser considerado uma extensão do modelo ARIMA. Modelos para memória longa na média foram introduzidos por Granger, C. W. J. e Joyeux, R. (1980) e Hosking, J. R. M. (1981), a partir do trabalho pioneiro de Hurst, H. E. (1951). Assim, um modelo ARFIMA incorpora o comportamento de memória longa ao introduzir a possibilidade de d ser fracionário. Pode ser escrito como

$$\Phi(B)(1-B)^d r_t = \Theta(B)a_t, \quad d \in \mathbb{R}, \quad (4.31)$$

onde, como antes, os polinômios $\Phi(B)$ e $\Theta(B)$ possuem ordens p e q , respectivamente. O processo $\{a_t\}_{t \in \mathbb{Z}}$ é um ruído branco com média zero e variância finita σ_a^2 . O termo $(1-B)^d$ é a expansão em séries binomiais dada por $(1-B)^d = 1 - dB + \frac{d(d-1)B^2}{2!} - \frac{d(d-1)(d-2)B^3}{3!} + \dots$ (veja propriedades em Diebold, F. X e Rudebusch, G. D. (1989) ou Hosking, J. R. M. (1981)).

Por causa de algumas suposições de modelos de mercados eficientes, séries de preços são em geral $I(1)$. Contudo, a primeira diferença dessas séries, mesmo se mostrando estacionárias, muitas vezes exibem uma correlação não nula entre observações distantes. Sua f.a.c. pode decair lentamente à uma taxa polinomial. A Figura 4.14 mostra a f.a.c. de séries supostamente $I(1)$ à esquerda, e $I(0)$ à direita. Na linha superior temos série de um índice americano, e na de baixo a série do índice IGPAGEN do Chile. A série “ $I(0)$ ” do Chile apresenta comportamento diferente da série “ $I(0)$ ” do mercado americano. Todas as autocorrelações são positivas e o decaimento é lento, caracterizando a existência de memória longa.

Este fenômeno é conhecido em Hidrologia, onde se observa a “persistência” nos níveis dos rios, conhecido como efeito “Hurst”. Além disto, estudos de Mandelbrot, B. B. (1963), Mandelbrot, B. B. (1969b), Mandelbrot, B. B. (1972) sobre o movimento Browniano fracionário levou à extensão dos modelos ARIMA, destinada a modelar persistência de longo termo.

Se $r_t = (1-B)^d p_t$, e d não é inteiro, r_t é dita ser fracionalmente integrada. Então os modelos adequados são os modelos ARFIMA, noção na literatura desde Hosking, J. R. M. (1981). Para $d > -1$,

$$\Delta^d = (1-B)^d$$

E a partir da expansão em séries binomiais temos, para $d > -1$,

$$\Delta^d = (1-B)^d = 1 - dB + \frac{d(d-1)B^2}{2!} - \frac{d(d-1)(d-2)B^3}{3!} + \dots$$

Considere o processo ARFIMA(0, d , 0)

$$(1-B)^d r_t = a_t \quad -0.5 < d < 0.5,$$

um *ruído branco fracionário*, RBF, análogo no tempo discreto do movimento browniano fracionário. Naturalmente que para $d = 0$, r_t é um ruído branco e sua f.a.c. amostral declina imediatamente para zero. Notemos que para $d = 1$, r_t é um *passeio aleatório* (um AR(1) com raiz unitária) e portanto sua f.a.c. fica próxima de 1 para todos os *lags*. Para valores fracionários de d a f.a.c. declina hiperbolicamente a zero, isto é, $\rho_k = \Gamma k^{2d-1}$, onde Γ é a função Gama. Quando $d \geq 0.5$ a variância de r_t é infinita, e o processo é não estacionário, mas não chega a ser um “não estacionário” como um $I(1)$. É memória longa porque para k grande, ρ_k ainda é significativamente diferente de zero.

Algumas propriedades deste modelo são (veja Hosking, J. R. M. (1981)):

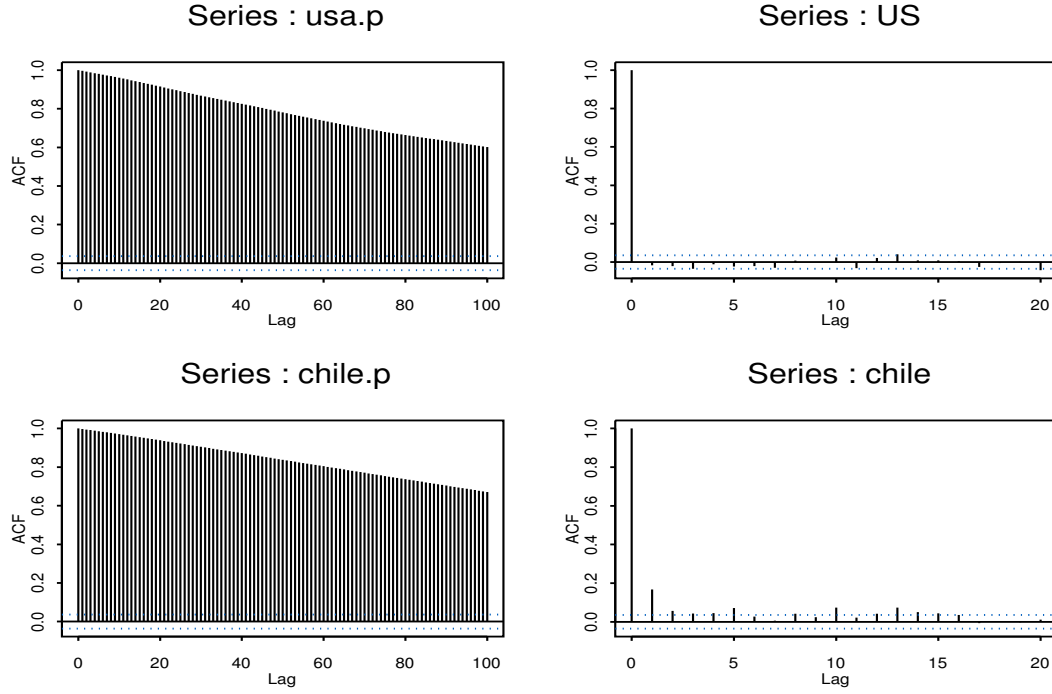


Figura 4.14: *f.a.c. de séries e $I(1)$ à esquerda e $I(0)$ ou $I(d)$, d fracionário, à direita.*

1. Se $d < 0.5$, então r_t é um processo fracamente estacionário e possui uma representação $MA(\infty)$:

$$r_t = a_t + \sum_{i=1}^{\infty} \psi_i a_{t-i}$$

$$\text{com } \psi_k = \frac{d(1+d) \cdots (k-1+d)}{k!} = \frac{(k+d-1)!}{k!(d-1)!}.$$

2. Se $d > -0.5$, então r_t é um processo inversível, e possui uma representação $AR(\infty)$:

$$r_t = a_t + \sum_{i=1}^{\infty} \pi_i r_{t-i} + a_t$$

$$\text{com } \pi_k = \frac{-d(d-1) \cdots (k-1-d)}{k!} = \frac{(k-d-1)!}{k!(-d-1)!}.$$

3. Se $-0.5 < d < 0.5$, então a f.a.c. de r_t é

$$\rho_k = \frac{d(1+d) \cdots (k-1+d)}{(1-d)(2-d) \cdots (k-d)}, \quad k = 1, 2, \dots,$$

e em particular $\rho_1 = \frac{d}{(1-d)}$ e $\rho_k \approx \frac{(-d)!k^{2d-1}}{(d-1)!}$, quando $k \rightarrow \infty$.

4. Se $-0.5 < d < 0.5$, então a f.a.c. parcial de r_t é

$$\phi_{k,k} = \frac{d}{k-d}, k = 1, 2, \dots$$

5. Se $-0.5 < d < 0.5$, então a função densidade espectral $f(w)$ de r_t , a qual é a transformada de Fourier da f.a.c. de r_t , satisfaz

$$f(w) \approx w^{-2d}, \text{ quando } w \rightarrow 0,$$

onde w é a frequência $w \in [0, 2\pi]$.

Assim, um modelo ARFIMA incorpora o comportamento de memória longa ao introduzir a possibilidade de d ser fracionário.

A intuição por detrás deste conceito de memória longa e sua ligação com o fato ou a restrição de ser $d < 1$, pode ser vista mais claramente no domínio da frequência. Para várias séries financeiras a densidade espectral (função potência) cresce quando a frequência se aproxima de zero. Muitas vezes o espectro parece ser infinito perto de zero o que poderia indicar ser a primeira diferenciação necessária e indicada. Contudo, a série diferenciada pode apresentar nenhum “poder” perto de zero, indicando que a primeira diferenciação foi excessiva. A Figura 4.15 ilustra este fato para a série de log-retornos do Chile. Lembremos que frequência perto de zero significa períodos grandes (exceto ciclos e/ou periodicidades, pois a série é estacionária).

Sugerimos gerar observações de um processo ruído branco fracionário com, por exemplo, $d = 0.3$, e obter a estimativa de sua densidade espectral baseada no periodograma suavizado, e observar as figuras.

O necessidade de se fazer ou não uma diferenciação poderia ser testada utilizando-se algum teste de raiz unitária. Esses testes, contudo, possuem potência (probabilidade de rejeitar a hipótese nula quando a mesma é falsa) ainda menor contra alternativas fracionárias. O melhor é aplicar testes específicos para detectar a existência de d fracionário.

A quantidade mais utilizada para testar a presença de memória longa é a razão entre o *range* e o desvio padrão, $\frac{R}{S}$, originalmente proposta em Mandelbrot, B. B. (1969a) e Mandelbrot, B. B. (1969b), no contexto da economia. Esta quantidade é estimada com a estatística RS , definida como a como a razão entre os estimadores \hat{R} e $\hat{\sigma}$, onde $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$, e onde

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{i=1}^T (r_i - \bar{r})^2, \quad \bar{r} = \frac{1}{T} \sum_{i=1}^T r_i,$$

e

$$\hat{R} = \max_{1 \leq i \leq T} \sum_{t=1}^i (r_t - \bar{r}) - \min_{1 \leq i \leq T} \sum_{t=1}^i (r_t - \bar{r}).$$

Note que \hat{R} é definida a partir de somas parciais, sendo as primeiras todas não negativas, e as seguintes todas sempre não positivas, o que garante uma estatística \hat{R} sempre não negativa.

A hipótese nula a ser testada é “ H_0 : Não existe memória longa”. Porém a estatística RS é sensível à presença de dependência de memória curta, k pequeno. Isto é, a presença de ρ_k significativo, para k pequeno, altera a distribuição de RS sob a hipótese nula, e pode levar a conclusões errôneas.

Lo, A. (1991) propôs a estatística RS modificada, \widehat{R}_q , que incorpora a presença de correlação de memória curta. Foi proposto o estimador $\widehat{R}_q = \widehat{R}/S_q$, onde $S_q = \sqrt{S_q^2}$, e onde

$$S_q^2 = S^2 \left(1 + \frac{2}{T} \sum_{j=1}^q w_{qj} \rho_{j,T} \right),$$

onde $w_{qj} = 1 - \frac{j}{q+1}$, $q < T$, onde $\rho_{j,T}$, $j = 1, 2, \dots, q$ são as autocorrelações amostrais de $\text{lag } j$.

Lo, A. (1991) obteve a distribuição assintótica de \widehat{R}_q , isto é, mostrou que $\frac{1}{\sqrt{T}} \widehat{R}_q$ converge em distribuição para uma v.a. bem definida, fornecendo os valores tabelados para o teste. A estatística \widehat{R}_q é consistente quando as alternativas consideradas são $-0.5 \leq d \leq 0.5$. Notemos que q é a ordem do MA no modelo ARFIMA(p, d, q), e que q deve ser escolhido antes de se efetuar o teste, problema que ainda não está muito bem resolvido.

Se r_t possuir caudas pesadas, a distribuição assintótica de \widehat{R}_q fica alterada (deslocada para esquerda), e os valores tabelados não servem mais como valores críticos. Isto nos faria rejeitar quando *não* deveríamos rejeitar na cauda esquerda ($d < 0$), mas *não* rejeitar quando deveríamos rejeitar na cauda direita. Lô sugere que este teste deva ser usado como *mais* uma ferramenta e não como a ferramenta decisória. Também recomenda usar $q = T^{1/4}$.

Aplicamos o teste \widehat{R}_q aos dados de log-retornos dos índices aqui utilizados para ilustração mas o teste não rejeitou a hipótese nula para nenhum deles. Para os dados de log-retornos da China o obtivemos o seguinte resultado:

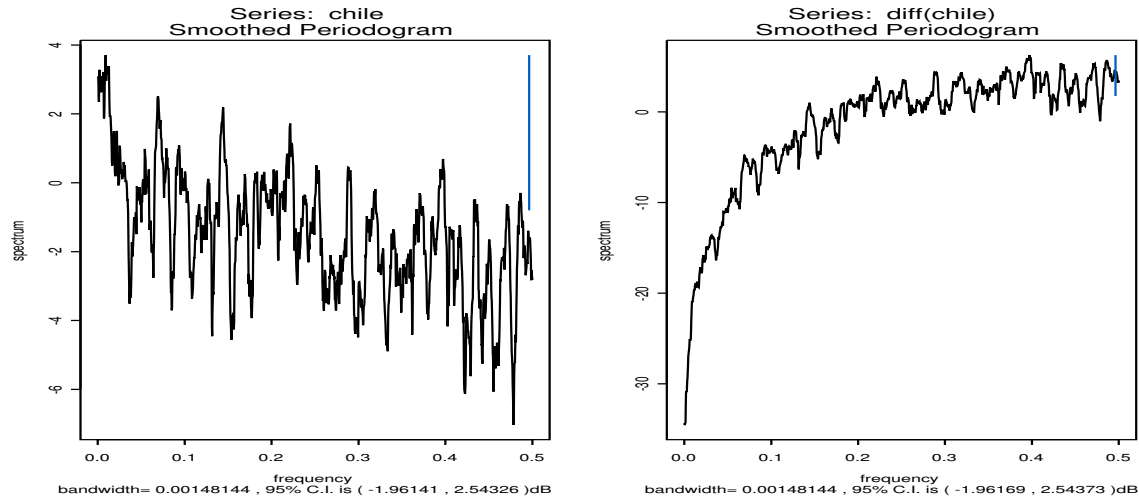


Figura 4.15: Periodograma da série log-retornos da China à esquerda, e de sua diferenciação, à direita.

Test Statistics: R/S

2.7117** ** : significant at 1% level

Test Statistics: Modified R/S Test (q=9, default)

2.1262** ** : significant at 1% level

Test Statistics: Modified R/S Test (q=7, L6)

2.1801** ** : significant at 1% level

A estimação de d pode ser feita por vários métodos, inclusive alguns baseados na densidade espectral, sendo os mais conhecidos aqueles baseados no modelo de regressão, os estimadores semi-paramétricos clássicos e robustos (veja Mendes, B. V. M. e Lopes, S. R. C. (2006)). Também pode-se obter uma estimativa para d a partir da estimação (global) do modelo ARFIMA(p, d, q). Por exemplo, a estimação do modelo ARFIMA(p, d, q) para a série de log-retornos da China, onde a otimização foi feita considerando como melhor modelo (dentre opções para p e q de 0 a 3) aquele minimizando o critério BIC, resultou em:

Coefficients:

	Value	Std. Error	t value	Pr(> t)
d	0.1329	0.0140	9.4808	0.0000

Information Criteria:

log-likelihood	BIC
-4180.436	8368.920

Residual scale estimate: 0.9216

BIC of all models estimated:

	q=0	q=1	q=2	q=3
p=0	8368.920	8374.155	8381.660	8387.787
p=1	8376.698	8383.566	8384.108	8392.416
p=2	8385.248	8391.070	8398.062	8401.270
p=3	8394.790	8401.574	8404.941	8412.979

Aceita hipótese de existencia de d-fracionario. P-valor: 0"

Então a série de preços da China é um $I(1.1329)$ em vez de $I(1)$. Mais especificamente, a série de log-retornos da China é um ruído branco fracionário, RBF .

4.5 Processos Não Estacionários

Os modelos estudados nas seções anteriores assumem que as séries são estacionárias, isto significando que as médias e as variâncias são constantes, e que as autocorrelações só dependem do *lag*. Entretanto,

séries financeiras tais como taxas de câmbio ou o preço de uma ação podem não ser estacionárias. Por exemplo, observe na Figura 1.1 as séries de preços diários dos quatro índices utilizados neste texto, no período de 03/01/2006 a 13/10/2011.

A não estacionariedade na média pode ser eliminada diferenciando-se a série. Alternativamente, podemos ajustar um polinômio, isto é, assumir que $\mu_t = \mu(t)$ e modelar como $\sum_{j=0}^k \beta_j t^j$. Na prática, é melhor diferenciar quantas vezes forem necessárias e depois modelar a série resultante com um modelo ARMA(p, q) ou ARFIMA(p, d, q).

Quando a não estacionariedade é na variância (não condicional), podemos fazer alguma transformação. Suponha que $r_t = \mu_t + a_t$, onde μ_t é não estocástica, e a_t é uma *v.a.*

Suponha que a variância do erro dependa do “nível médio” da série μ_t , isto é

$$\text{var}(r_t) = \text{var}(a_t) = \sigma^2 h^2(\mu_t)$$

onde $h(\cdot)$ é uma função conhecida. Neste caso, a variância não é constante, dependendo de t através de μ_t . Esta forma de não estacionariedade pode ser corrigida se acharmos uma função $g(\cdot)$ tal que $g(r_t)$ estabilize a variância. Isto é, devemos ter

$$\text{var}(y_t) = \text{var}(g(r_t)) = \text{constante}.$$

Para achar g podemos expandir $g(r_t)$ em série de Taylor em torno de μ_t :

$$g(r_t) \approx g(\mu_t) + (r_t - \mu_t)g'(\mu_t)$$

$$\begin{aligned} \text{var}(g(r_t)) &\approx [g'(\mu_t)]^2 \text{var}(r_t) \\ &= [g'(\mu_t)]^2 h^2(\mu_t) \sigma^2. \end{aligned}$$

E então, para estabilizar a variância, tome $g(\cdot)$ tal que $g'(\mu_t) = \frac{1}{h(\mu_t)}$.

Por exemplo, suponha que para uma série, quanto maior a média, maior o desvio padrão. Neste caso μ_t é proporcional a σ e poderíamos ter $h(\mu_t) = \mu_t$. Assim, $g'(\mu_t) = \frac{1}{\mu_t}$, e portanto $g(\mu_t) = \log(\mu_t)$. Outra transformação muito adequada é $2\sqrt{\mu_t} = g(\mu_t)$. Essas duas transformações são bastante úteis em finanças, as quais são, na verdade, casos particulares da transformação Box-Cox.

4.6 Testes de Raiz Unitária

A não estacionariedade de uma série de preços de ativos financeiros está quase sempre ligada ao fato de não existir um nível fixo, comum para as observações. Veja novamente a Figura 1.1. Séries temporais apresentando este tipo de não estacionariedade são chamadas de *séries temporais não estacionárias com raiz unitária*. O exemplo mais famoso de série temporal não estacionária com raiz unitária é o PA.

Conforme vimos na seção 4.2, o passeio aleatório, PA, é o processo estocástico p_t tal que $p_t = p_{t-1} + a_t$, com $a_t \sim \text{RB}$. É interessante ver o PA como um AR(1) com coeficiente $\phi = 1$:

$$p_t = \phi p_{t-1} + a_t.$$

Vemos que se a_t tiver distribuição simétrica em torno de zero, condicionalmente a p_{t-1} , p_t será maior que p_{t-1} com probabilidade $1/2$, e menor que p_{t-1} com probabilidade $1/2$. Ora, $\phi = 1$ não satisfaz a condição de estacionaridade fraca, sendo portanto o PA um exemplo de série temporal não estacionária com raiz unitária. Interessante notar como são as previsões baseadas no PA:

$$\widehat{p_h(1)} = E[p_{h+1}|p_h, p_{h-1}, \dots] = E[p_h + a_{h+1}|\mathcal{L}_h] = p_h.$$

A previsão 2-passos-a-frente é dada por

$$\widehat{p_h(2)} = E[p_{h+2}|p_h, p_{h-1}, \dots] = E[p_{h+1} + a_{h+2}|\mathcal{L}_h] = \widehat{p_h(1)} = p_h,$$

e assim por diante. Portanto temos $\widehat{p_h(l)} = p_h$, sem utilidade, não sendo do tipo que reverte para a média.

Para entender melhor o processo PA, vejamos sua representação como um MA.

$$\begin{aligned} p_t &= p_{t-1} + a_t \\ &= p_{t-2} + a_{t-1} + a_t = p_{t-3} + a_{t-2} + a_{t-1} + a_t. \\ p_t &= a_t + a_{t-1} + a_{t-2} + \dots \end{aligned}$$

um $\text{MA}(\infty)$. Já sabemos que a previsão l -passos-a-frente é p_h , e sendo $p_h = a_h + a_{h-1} + a_{h-2} + \dots$, o erro de previsão é

$$\begin{aligned} e_h(l) &= p_{h+l} - \widehat{p_h(l)} \\ &= a_{h+l} + a_{h+l-1} + a_{h+l-2} + \dots + a_{h+1} + a_h + a_{h-1} + \dots - a_h - a_{h-1} - \dots \\ e_h(l) &= a_{h+l} + a_{h+l-1} + a_{h+l-2} + \dots + a_{h+l}, \end{aligned}$$

e a variância do erro de previsão é

$$\text{var}(e_h(l)) = l\sigma_a^2,$$

a qual diverge para ∞ quando $l \uparrow \infty$, e como consequência o tamanho do IC da previsão cresce para ∞ , não sendo prático. Em outras palavras, o modelo PA não é previsível. Além disto, a variância não condicional de p_t não é limitada (veja $\text{var}(\text{MA}(\infty))$), o que quer dizer que no futuro o preço p_t pode ser qualquer coisa, ou muito grande ou muito pequeno. Neste sentido, vemos que o PA não é uma boa representação para séries de preços. Para completar, notemos que a representação do PA como um $\text{MA}(\infty)$, $p_t = a_t + a_{t-1} + a_{t-2} + \dots$ nos diz que o choque corrente e todos os choques passados têm o mesmo peso, ou contribuem da mesma forma. Neste caso, dizemos que a série temporal tem uma memória *forte* e lembra de todos os choques passados. Em Economia se diz que “os choques tem efeito permanente” na série.

PA com nível. Os retornos financeiros costumam ter μ pequeno e positivo. Podemos modificar o PA para refletir esta característica:

$$p_t = \mu + p_{t-1} + a_t$$

sendo $a_t \sim \text{RB}(0, \sigma_a^2)$. Notemos então que $\mu = E[p_t - p_{t-1}]$, e portanto μ representa a tendência no tempo (o que se espera no longo prazo), chamado de *drift* do modelo. Para ver o efeito de μ , seja p_0 o valor inicial do preço. Teremos:

$$p_t = \mu + p_0 + a_1$$

$$\begin{aligned} p_2 &= \mu p_1 + p_2 + a_2 = \mu + \mu + p_0 + a_1 + a_2 \\ &= 2\mu + p_0 + a_1 + a_2 \end{aligned}$$

... e assim por diante, obtemos

$$p_t = t\mu + p_0 + a_1 + a_2 + \cdots + a_t$$

e temos que o log do preço é a soma de $t\mu$ com um PA puro, $\sum_{j=1}^t a_j$.

A variância de p_t é $t\sigma_a^2$ e o desvio padrão é $\sqrt{t}\sigma_a$. Compare com $E[p_t|\cdot] = t\mu + p_0$, e veja que o desvio padrão cresce à uma taxa menor que a esperança. O gráfico de $E[p_t|\cdot]$ versus o tempo t é uma reta com intercepto p_0 e inclinação μ , ilustrando a não estacionaridade na média. Assim, sendo p_t o log do preço, temos $r_t = p_t - p_{t-1}$ e portanto $r_t = \mu t + a_t - \mu(t-1)$, onde $r_t = \mu(t-t+1) + a_t$ e assim $r_t = \mu + a_t$, isto é, os log-retornos seriam um RB com esperança igual a μ , onde μ é a inclinação do PA com *drift*.

Como

$$E(p_t) = p_0 + t\mu$$

$$\text{var}(p_t) = t\sigma^2 = \gamma_{0,t}$$

$$\gamma_{k,t} = \text{cov}(p_t, p_{t-k}) = E[(a_1 + \cdots + a_t)(a_1 + \cdots + a_{t-k})] = E[a_1^2 + \cdots + a_{t-k}^2] = (t-k)\sigma^2, k \geq 0$$

temos que

$$\rho_{k,t} = \frac{(t-k)}{\sqrt{t(t-k)}} = \sqrt{\frac{t-k}{t}},$$

e notemos que se t é grande quando comparado com k , os $\rho_{k,t} \approx 1$.

Então, a medida que o tempo passa, as correlações para *lags* pequenos ficam ≈ 1 , a sequência dos p_t fica amortecida mas não estacionária, porque as médias e as variâncias continuam crescendo.

Na verdade, o PA é um exemplo de processo não estacionário pertencente à uma grande classe de processos não estacionários, conhecida como *processos integrados*.

Modelos gerais não estacionários de raiz unitária. Veja antes uma introdução a este assunto dada no Capítulo 1. Seja um modelo ARMA(p, q). Façamos agora uma das raízes características do polinômio AR ser igual a 1. Teremos um modelo ARIMA, um modelo não estacionários com raiz unitária. Um ARIMA, assim como um PA, tem memória forte, isto porque os coeficientes ψ_i na sua representação MA, não decaem a zero com o tempo. A abordagem mais utilizada para tratar a raiz unitária é diferenciar.

Uma série temporal y_t é uma ARIMA(p, d, q), d inteiro, se a série transformada

$$r_t = y_t - y_{t-1} = (1 - B)y_t \sim \text{ARMA}(p, q) \text{ estacionário, inversível.}$$

por exemplo, o log-preço $= y_t$ seria ARIMA($p, 1, q$). A série r_t é então a série diferenciada. Quando for necessário diferenciar mais vezes para obter uma série temporal estacionária, significa que a série teria várias raízes unitárias.

Exemplo: z_t tem 2 raízes unitárias. $y_t = z_t - z_{t-1}$ teria 1 raiz unitária e $r_t = y_t - y_{t-1} = z_t - 2z_{t-1} + z_{t-2}$ seria ARMA(p, q) estacionário. Neste caso z_t seria ARIMA($p, 2, q$), ou seja, z_t é integrado de ordem

2, um $I(2)$. Precisamos então saber testar a existência de raiz unitária. Para o desenvolvimento do teste de raiz unitária, usaremos o PA.

Suponha que queremos testar se p_t é PA, isto é, $p_t = \phi_1 p_{t-1} + a_t$ com $\phi_1 = 1$. As hipóteses são

$$H_0 : \phi_1 = 1 \quad vs \quad H_1 : \phi_1 < 1, \quad (4.32)$$

e podemos usar o teste DF. Dickey, D. A. e Fuller, W. A. (1979), abordaram este problema e propuseram a estatística teste DF, que teria distribuição t , apenas se $|\phi_1| < 1$, como numa regressão. Esta estatística (t -ratio se $|\phi_1| < 1$) é

$$DF = \frac{\hat{\phi}_1 - 1}{std(\hat{\phi}_1)},$$

onde $\hat{\phi}_1$ é o estimador de MQ de ϕ_1 .

Sabemos que

$$\hat{\phi}_1 = \frac{\sum_1^T p_t p_{t-1}}{\sum_1^T p_{t-1}^2} \quad e \quad \hat{\sigma}_a^2 = \frac{\sum_1^T (p_t - \hat{\phi}_1 p_{t-1})^2}{T - 1}$$

com $p_0 = 0$. Obtemos então $std(\hat{\phi}_1) = \frac{\hat{\sigma}_a}{\sqrt{\sum_1^T p_{t-1}^2}}$.

Logo,

$$DF = \frac{\hat{\phi}_1 - 1}{std(\hat{\phi}_1)} = \frac{\frac{\sum_1^T p_{t-1} p_t}{\sum_1^T p_{t-1}^2} - 1}{\frac{\hat{\sigma}_a}{\sqrt{\sum_1^T p_{t-1}^2}}} = \frac{\frac{\sum_1^T p_{t-1} p_t - \sum_1^T p_{t-1}^2}{\sum_1^T p_{t-1}^2}}{\frac{\hat{\sigma}_a}{\sqrt{\sum_1^T p_{t-1}^2}}} = \frac{\sum_1^T p_{t-1} a_t}{\hat{\sigma}_a \sqrt{\sum_1^T p_{t-1}^2}}$$

chamada estatística teste de Dickey-Fuller.

Se $\{a_t\} \sim RB$ com momentos de ordem maior que 2 finitos, então

$$DF \xrightarrow[t \rightarrow \infty]{} g(W)$$

onde g é alguma função e W representa o movimento Browniano. Temos o resultado

$$DF \xrightarrow[T \rightarrow \infty]{d} \frac{\frac{1}{2}([W(1)]^2 - 1)}{(\int_0^1 [W(r)]^2 dr)^{\frac{1}{2}}} \quad (4.33)$$

onde $W(r)$ é o movimento Browniano padrão, isto é, para cada t , $W(t) \sim N(0, t)$. Em particular $(W(1))^2 \sim \chi^2(1)$. A distribuição da estatística do lado direito de 4.33 foi obtida por Dickey, D. A. e Fuller, W. A. (1979). Se $\widehat{DF} < \text{valor crítico}$, rejeitamos H_0 (teste unilateral). Alguns valores críticos de DF são dados na Tabela 4.3.

Suponha agora que a média não seja zero no modelo PA, $p_t = \phi_1 p_{t-1} + a_t$, e temos o modelo

$$p_t = \theta_0 + \phi_1 p_{t-1} + a_t \quad (4.34)$$

Tabela 4.3: Valores críticos do teste DF.

	1%	5%	10%
	Valor Crítico		
T=100	-2.6	-1.95	-1.61
T > 500	-2.58	-1.95	-1.62

onde $a_t \sim \text{RB}(0, \sigma_a^2)$, $\theta_0 = (1 - \phi_1)\mu$, $\mu = E(p_t)$ e $|\phi_1| < 1$. Note que (4.34) é o mesmo que $p_t - \mu = \phi_1(p_{t-1} - \mu) + a_t$. Queremos testar se existe raiz unitária, hipóteses (4.32). Observe que subtraindo p_{t-1} nos 2 lados de 4.34, obtemos:

$$p_t - p_{t-1} = \theta_0 + \phi_1 p_{t-1} - p_{t-1} + a_t$$

$$\Delta p_t = \theta_0 + \phi^* p_{t-1} + a_t$$

onde $\phi^* = \phi_1 - 1$. Podemos obter o estimador de MQ de ϕ^* fazendo a regressão de Δp_t sobre p_{t-1} .

Notemos que as hipóteses (4.32) equivalem a

$$H_0 : \phi^* = 0 \quad \text{vs} \quad H_1 : \phi^* < 0$$

A estatística teste é $DF_\mu = \frac{\hat{\phi}_{MQ}^*}{\widehat{\text{e.p.}}(\hat{\phi}_{MQ}^*)}$, ver detalhes em Tsay (2003).

$$DF_\mu \xrightarrow{d} \frac{\frac{1}{2}([W(1)]^2 - 1) - W(1)\int_0^1 W(r)dr}{[\int_0^1 [W(r)]^2 dr - (\int_0^1 W(r)dr)^2]^{\frac{1}{2}}} \quad (4.35)$$

A distribuição desta estatística se afasta ainda mais da Normal. Este é o teste de DF aumentado ou ADF. Para amostras grandes ($T > 500$), valores críticos estão em Fuller, W. A. (1976, cap. 8).

Uma extensão simples do teste DF implementada em vários pacotes computacionais é a seguinte. Suponha agora que a série seja um $\text{AR}(p)$

$$r_t - \mu = \phi_1(r_{t-1} - \mu) + \dots + \phi_p(r_{t-p} - \mu) + a_t \quad (4.36)$$

onde $a_t \sim \text{RB}(0, \sigma_a^2)$, que pode ser reescrito como

$$r_t = \theta_0 + \sum_{i=1}^p \phi_i r_{t-i} + a_t \quad (4.37)$$

ou ainda como

$$\Delta r_t = \theta_0 + \phi_1^* r_{t-1} + \phi_2^* \Delta r_{t-1} + \dots + \phi_p^* \Delta r_{t-p+1} + a_t \quad (4.38)$$

onde $\phi_1^* = \sum_{i=1}^p \phi_i - 1$ e $\phi_j^* = -\sum_{i=j}^p \phi_i$, $j = 2, \dots, p$.

Se o polinômio autorregressivo $\phi(B)$ tiver uma raiz unitária, então $\phi(1) = 1 - \phi_1 - \phi_2 - \dots - \phi_p = 0$, ou seja, $\sum_{i=1}^p \phi_i = 1$ e $\phi_1^* = 0$. Portanto testar a existência de raiz unitária em um $\text{AR}(p)$ é equivalente a testar

$$H_0 : \phi_1^* = 0,$$

onde $\phi_1^* = \sum_{i=1}^p \phi_i - 1$. Segue que ϕ_1^* pode ser estimador MQ com o coeficiente da regressão onde a variável dependente é Δr_t , e as variáveis explicativas são $1, r_{t-1}, \Delta r_{t-1}, \dots, \Delta r_{t-p+1}$. A estatística teste tem distribuição assintótica igual a dada anteriormente.

Notemos que se $r_t \sim \text{ARMA}(p, q)$, também podemos usar este teste. Vale a pena lembrar que neste processo de estimação, a ordem p é obtida na medida que se determina quantos coeficientes da regressão são estatisticamente significativos. Os resíduos desta regressão devem aceitar a hipótese nula do teste de Ljung-Box. Uma outra alternativa para a determinação da ordem de integração de uma série é dada pelo teste de Phillips-Perron.

4.7 Exercícios

Exercício 1: Obtenha uma série de preços de uma ação brasileira e a partir dela obtenha a série de log-retornos $\{r_t\}$. Verifique se esta série é estacionária. Aplique os testes RS e de raiz unitária.

Exercício 2: Para a série acima ajuste o melhor modelo da classe ARFIMA(p, d, q).

Exercício 3: Suponha uma série de retornos com $T = 1000$. Supondo que os valores estimados para os $\phi_{j,j}$ sejam aqueles dados na tabela abaixo, escolha uma ordem adequada p para ajuste de AR(p).

Tabela 4.4: Estimativas dos coeficientes de autocorrelação parcial, $\hat{\phi}_{j,j}$, para $j = 1, \dots, 10$.

p	1	2	3	4	5	6	7	8	9	10
f.a.c.p.	0.19	-0.16	-0.11	0.03	0.007	-0.05	0.02	0.06	0.07	-0.12

Capítulo 5

Volatilidade

Volatilidade não é a mesma coisa que risco. Como uma medida de incerteza, tem papel fundamental na alocação de ativos, no apreçamento de opções, na administração do risco, em técnicas de *hedging* e na política monetária. Esta abrangência e importância faz com que muito da atividade na área de pesquisa de uma instituição financeira esteja voltada para a modelagem e previsão da volatilidade de ativos financeiros. É tão importante no apreçamento de derivativos, que o mercado lista o preço de opções em termos de unidades de volatilidade. Pode-se também comprar derivativos escritos como volatilidade. Assim, a previsão da volatilidade, ou da volatilidade da volatilidade, se torna necessária e de suma importância. Na administração de carteiras, uma boa previsão da volatilidade dos preços dos ativos durante o período de investimento é o ponto de partida para se acessar o risco do investimento. A volatilidade, não sendo observável, deve ser estimada, e por isto sofre de várias formas de incerteza, incluindo na especificação do modelo e na estimação dos seus parâmetros.

Desde o primeiro acordo da Basileia em 1996, a administração do risco tornou-se central em qualquer instituição financeira. E isto fez com que a estimação e previsão da volatilidade se tornasse a chave deste processo. Muitas vezes uma estimativa da volatilidade chega a balizar e medir a vulnerabilidade de um mercado financeiro e de uma economia.

Talvez a maneira mais simples de se medir a volatilidade não condicional de um conjunto de dados, seja através do desvio padrão amostral S , definido no Capítulo 1, as vezes chamado de volatilidade histórica. Apesar de poder ser calculado para qualquer distribuição, somente sob a suposição de normalidade a distribuição de S^2 é conhecida analiticamente. Nos outros casos, a distribuição de S^2 pode ser obtida por simulação. Dada uma série de retornos, S pode ser calculado utilizando-se as últimas N observações (janela de tamanho N), e deslocar esta janela ao longo do tempo. Notemos que sob normalidade das observações S^2 é um estimador não viciado de σ^2 , mas S não o é de σ .

Estendendo a idéia de volatilidade histórica, temos vários outros métodos: média móvel, *Exponential Smoothing*, e o EWMA, *exponentially weighted moving average*, o *smooth transition exponential smoothing* de James Taylor (2001), entre outros. O *RiskmetricsTM* usa o método EWMA. Outra possibilidade é usar $|r_t|$ ou r_t^2 .

Para se estimar a volatilidade condicional um modelo deve ser assumido. Em um modelo linear, o valor do processo r_t no instante t depende de uma função linear do valor presente e dos valores passados da inovação a_t . Se um modelo é não-linear, supomos que as inovações a_t são i.i.d. e o modelo é da forma

$$r_t = g(a_{t-1}, a_{t-2}, \dots) + a_t h(a_{t-1}, a_{t-2}, \dots),$$

para algumas funções g e h . Desta forma, $g(\cdot)$ representa a média condicional, e $h^2(\cdot)$ a variância condicional. Se $g(\cdot)$ for não linear, o modelo será não linear na média. Se $h(\cdot)$ for não linear, o modelo será não linear na variância. Por exemplo, o modelo $r_t = a_t + \alpha a_{t-1}^2$ é não-linear na média, pois $h(\cdot) \equiv 1$ e $g(\cdot) = \alpha a_{t-1}^2$. Considere agora o modelo $r_t = a_t \sqrt{\alpha r_{t-1}^2}$. Este é não-linear na variância, pois $g(\cdot) \equiv 0$ e $h(\cdot) = \sqrt{\alpha r_{t-1}^2}$ e assim r_{t-1} depende de a_{t-1} . Note que $\text{var}(r_t | \mathcal{I}_{t-1}) = \alpha r_{t-1}^2 = h^2(\cdot)$. Este é o modelo *ARCH(1)*. Outros exemplos de modelos não lineares:

(a) *Modelos Polinomiais*. Neste modelo temos

$$g(a_{t-1}, a_{t-2}, \dots) = \sum_{i=1}^{\infty} c_i a_{t-i} + \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} b_{ij} a_{t-i} a_{t-j} + \dots$$

que envolve termos lineares, bilineares, etc, das inovações.

(b) *Modelos Bilineares*. Neste modelo temos

$$g(a_{t-1}, a_{t-2}, \dots) = \sum_{i=1}^{\infty} \alpha_i a_{t-i} + \sum_{i=1}^{\infty} \beta_i r_{t-i} + \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \gamma_{ij} r_{t-i} a_{t-j}.$$

Na Seção 5.2 veremos alguns modelos da família GARCH para a volatilidade condicional. Na Seção 5.3 veremos um modelo de volatilidade estocástica com memória longa e na Seção 5.4 falaremos um pouco sobre medidas *realizadas* de variabilidade obtidas a partir de dados de alta frequência. Estudos recentes, por exemplo Andersen, T. G., Bollerslev, T., Diebold, F. X. e Labys, P. (2003), têm sugerido que modelos simples baseados na volatilidade realizada superam os populares modelos GARCH ou os modelos de volatilidade estocástica para previsões fora da amostra. Melhores resultados podem ser ainda obtidos se incluirmos descontinuidades na modelagem, através de *jumps* (Andersen, T. G., Bollerslev, T. e Diebold, F. X. (2007), Maheu, J. M. e McCurdy, T. H. (2004)).

5.1 Fatos estilizados sobre a volatilidade condicional

A estimativa da volatilidade de retornos financeiros fica mais precisa quando se usa modelos que atribuem peso maior às informações mais recentes. Notemos que a prática consagrou a palavra *volatilidade* para descrever a variância condicional σ_t^2 de uma variável. Na administração de riscos, a volatilidade tem papel importante pois permite calcular medidas de risco condicionais, em particular o Valor-em-Risco (Capítulo 7) de uma certa posição condicional à situação corrente da economia. Estimativas para a volatilidade de r_t são requeridas em vários modelos em finanças e também no cálculo de várias medidas de risco. Um exemplo importante é na fórmula de Black-Scholes para precificação de opções. Nela, σ_t é o desvio padrão condicional do retorno r_t de alguma ação. Note que ela depende do tempo t e sua expressão é derivada sob a suposição de normalidade.

A volatilidade não pode ser medida diretamente, mas se caracteriza por: (i) aparecer em grupos de maior ou menor volatilidade, ou seja, períodos onde os retornos apresentam uma maior ou menor amplitude (Mandelbrot (1963), Fama (1965), Baillie et al. (1996)); (ii) evoluir continuamente no tempo, sendo considerada estacionária; (iii) reagir diferentemente a valores positivos ou negativos da série de retornos; (iv) apresentar persistência alta e reversão para a média; (v) ser também parcialmente explicada por variáveis exógenas.

A idéia básica por detrás do estudo da volatilidade é a de que séries (estacionárias) de retornos financeiros $\{r_t\}$ são em geral *não autocorrelacionadas* mas são *dependentes*. Veja novamente a Figura 1.3 onde a f.a.c. amostral dos quadrados dos retornos claramente sugere que os retornos não são independentes.

Dado o conjunto I_{t-1} de informações passadas até o tempo $t-1$, a esperança condicional $\mu_t = E[r_t | \mathcal{I}_{t-1}]$ pode ser especificada, como já vimos, por um modelo simples, um ARFIMA(p, d, q). Nada impede que a equação para a média condicional inclua também algumas variáveis explicativas através do modelo de regressão com erros tipo série temporal. A volatilidade é dada por $\sigma_t^2 \equiv \text{var}(r_t | \mathcal{I}_{t-1})$. A média não condicional é $\mu \equiv E[r_t] = E[E[r_t | I_{t-1}]]$. A variância não condicional é $\sigma^2 \equiv \gamma_0 \equiv \text{var}(r_t) = E[(r_t - \mu)^2]$. Neste capítulo a distribuição condicional de r_t será denotada por F e a distribuição não condicional de $\frac{r_t - \mu}{\sigma}$ será representada por H .

O terceiro (assimetria) e o quarto (curtose) momentos não condicionais são dados por

$$A = \frac{E[(r_t - \mu)^3]}{\sigma^3} \quad C = \frac{E[(r_t - \mu)^4]}{\sigma^4}$$

e a assimetria e curtose condicionais são analogamente dadas por

$$A_t = \frac{E[(r_t - \mu_t)^3 | \mathcal{I}_{t-1}]}{\sigma_t^3} \quad C_t = \frac{E[(r_t - \mu_t)^4 | \mathcal{I}_{t-1}]}{\sigma_t^4}.$$

Definimos a variação proporcional da variância condicional, ou o *retorno da variância* condicional como

$$rv_t \equiv \frac{\sigma_t^2 - \sigma_{t-1}^2}{\sigma_{t-1}^2}. \quad (5.1)$$

O retorno da variância condicional, rv_t , pode ser modelado como a soma de dois componentes, um previsível e um choque estocástico. A *volatilidade da volatilidade*, VoV , é então o desvio padrão deste componente aleatório, isto é

$$VoV \equiv \sqrt{\text{var}(rv_t)}.$$

A partir de um modelo para a volatilidade podemos também obter a *estrutura a termo da volatilidade*. Seja $\sigma_{t+k|t}^2 \equiv E[r_{t+k}^2 | \mathcal{I}_t]$. A estrutura a termo da volatilidade é a previsão $\sigma_{t+k|t}$ do desvio padrão para várias maturidades k , todas iniciando na data t . Assim, para um ativo com maturidade no tempo $t+k$, ela é definida como

$$v_{t+k|t} \equiv \sqrt{\text{var}_t\left(\sum_{j=1}^k r_{t+j}\right)} \cong \sqrt{\sum_{j=1}^k \sigma_{t+j|t}^2}$$

o que é aproximadamente a raiz quadrada da soma das próximas k previsões da variância condicional do retorno, estando no tempo t , isto é,

$$v_{t+k|t} \cong \sqrt{\sum_{j=1}^k E[r_{t+j}^2 | \mathcal{I}_t]}.$$

A estrutura a termo da volatilidade resume as propriedades das previsões feitas para o segundo momento do retorno. A partir destas previsões é possível definir várias características do processo da volatilidade. Mais a frente iremos calcular essas quantidades para algumas séries de ativos.

Conforme já salientamos, retornos financeiros exibem *clusters* de volatilidade. Isto significa que choques correntes na volatilidade afetam volatilidades futuras (assim, um bom modelo de volatilidade deve ser capaz de capturar este fato estilizado). Em outras palavras, a volatilidade é *persistente*.

Considere agora que a previsão $\sigma_{t+k|t}^2$ seja dada pela função de perda quadrática, isto é,

$$\sigma_{t+k|t}^2 \equiv E[(r_{t+k} - \mu_{t+k})^2 | \mathcal{I}_t],$$

que significa que a previsão da volatilidade k passos a frente depende do conjunto de informações disponíveis no tempo t , em particular do retorno no tempo t . A volatilidade é dita persistente se o retorno hoje, r_t , tem grande efeito na previsão da variância condicional vários passos a frente. Tomando a derivada de $\sigma_{t+k|t}^2$ em relação a r_t^2 temos um número (já que ambos estão na mesma unidade), denominado *forward persistence*:

$$\theta_{t+k|t} = \frac{\partial \sigma_{t+k|t}^2}{\partial r_t^2}. \quad (5.2)$$

A análise de $\theta_{t+k|t}$ nos informa sobre a persistência da volatilidade, o quanto a informação de hoje é relevante para a futura variância condicional. Para muitos modelos esta quantidade, apesar de decrescer geométricamente, pode ainda ser importante até mesmo um ano a frente.

A persistência acumulada é definida como

$$\begin{aligned} \phi_{t+k|t} &= \frac{\partial(\frac{1}{k}(\sigma_{t+k|t}^2 + \sigma_{t+k-1|t}^2 + \cdots + \sigma_{t+1|t}^2))}{\partial r_t^2} \\ &= \frac{1}{k}(\theta_{t+k|t} + \theta_{t+k-1|t} + \cdots + \theta_{t+1|t}) \end{aligned} \quad (5.3)$$

que definida como o efeito do retorno corrente na média das previsões das variâncias condicionais nos próximos períodos $t+1, \dots, t+k$, é dada pela média das persistências para vários passos a frente. Por exemplo, a reação dos preços de opções de longo prazo a choques da volatilidade indicam que um bom modelo de volatilidade deveria produzir estimativas com persistência acumulativa de, digamos, um ano, ainda bastante significativa (Engle, R. F. e Patton, A. J., 2001).

Uma outra medida de persistência é a *meia vida* da volatilidade definida como

$$\tau = k : |\sigma_{t+k|t}^2 - \sigma^2| = \frac{1}{2} |\sigma_{t+1|t}^2 - \sigma^2| \quad (5.4)$$

e que representa, a partir de t , o tempo necessário para que a volatilidade retorne metade do valor com o qual, em t , se afastou do nível médio ou variância não condicional. Naturalmente, quanto maior a persistência, maior τ .

A propriedade de reversão para a média da volatilidade indica que deve existir um nível *normal* de volatilidade ao qual a mesma retornaria depois de um período de alta (ou baixa) volatilidade. Notemos que este nível médio pode não ser constante ao longo do tempo e também depender do ambiente econômico. Esta propriedade implica que $\theta_{t+k|t}$ converge em probabilidade para zero, para todo t , isto é, a informação corrente não tem efeito na previsão de longo prazo, o que também significa que o limite em probabilidade de $\sigma_{t+k|t}^2$ é σ^2 , para $\sigma^2 < \infty$.

Outras variáveis podem conter informações relevantes para a série da volatilidade. Isto é bem natural já preços de ativos evoluem sofrendo influências dos mercados local e global. Existem na literatura inúmeros trabalhos propondo diversas séries diferentes a serem consideradas como explicativas, inclusive séries de medidas de variabilidade realizadas (veja em Accioly, V. B. e Mendes, B. V. M. (2015)). Além de séries econômicas e financeiras, eventos isolados também podem influenciar a volatilidade. Esses incluem resultados macro-econômicos, das próprias firmas, dia da semana, etc.

Finalmente, não existe uma boa solução para o problema da avaliação da qualidade de uma previsão da volatilidade, já que ela não é observável. Muitos autores simplesmente dividem cada retorno pela previsão um passo a frente do seu desvio padrão e testam se o quadrado desta série ainda seria previsível. Um outro método bastante utilizado é o de se ajustar um modelo de regressão linear simples onde a variável dependente seria a série de retornos ao quadrado e a variável explicativa a variância condicional um passo a frente. Deve-se testar se o intercepto é zero e se a inclinação é um. Este método permite a comparação de vários modelos de previsão da volatilidade. Contudo, este método *não* é recomendado devido ao fato de que a série dos quadrados dos retornos ser heteroscedástica, implicando em estimativas não eficientes dos coeficientes da regressão. Além disto, r_t^2 como um estimador da variância tem uma grande variabilidade (*noisy estimator*), o que implicará em um coeficiente de explicação R^2 da regressão muito pequeno. Alternativamente, podemos substituir os retornos ao quadrado por alguma medida de variância realizada.

5.2 Modelos GARCH

Séries de retornos financeiros apresentam várias formas de dinâmicas lineares e não-lineares, a mais forte sendo a dependência de sua variabilidade instantânea no seu próprio passado. Devido a essas e outras características bastante conhecidas de séries financeiras, Engle, R. F. (1982) introduziu os modelos autorregressivos condicionalmente heteroscedásticos (ARCH), cujo mérito é modelar os *clusters* de volatilidade destas séries. Tais modelos foram generalizados em Bollerslev, T. (1986) que introduziu os modelos generalizados autorregressivos condicionalmente heteroscedásticos (GARCH).

Entre as diversas abordagens para a modelagem da volatilidade, aquela através de modelos da família GARCH é inequivocamente a mais popular, isto provavelmente devido à flexibilidade desses modelos, facilidade de estimação e, conforme já dissemos, pela sua capacidade de reproduzir alguns fatos estilizados apresentados pela maioria destas séries, incluindo o excesso de curtose na distribuição

não condicional dos dados e heteroscedasticidade condicional que dão origem aos *clusters* de volatilidade. Após o artigo de Engle de 1982 inúmeros outros se seguiram com aplicações e extensões.

Modelo ARCH(m)

O modelo ARCH(m) pode ser escrito como

$$r_t = \mu_t + \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \dots + \alpha_m r_{t-m}^2,$$

onde $\{\varepsilon_t\}$ é uma sequência de v.a.'s i.i.d. com média zero e variância 1, isto é, $\varepsilon_t \sim \text{i.i.d. } F(0, 1)$, $\alpha_0 > 0$ e $\alpha_i \geq 0$ para $i = 1, \dots, m$. A soma dos parâmetros α_i , $\sum_{i=1}^m \alpha_i$, mede a persistência e deve ser menor que um para assegurar que a variância não condicional de r_t seja finita. Na prática, em geral supomos F sendo a distribuição Normal ou t-student com ν graus de liberdade (t_ν). Observe que pela definição, valores grandes de $|r_t|$ tendem a ser seguidos por outros valores grandes. Mesmo assumindo F como sendo a distribuição Normal pode-se mostrar que r_t tem excesso de curtose positivo, isto é, a distribuição não condicional dos retornos tem caudas pesadas. Este fato também explica os pontos extremos e a não normalidade de r_t .

Considere agora o processo ARCH(1): $r_t = \sigma_t \varepsilon_t$, $\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2$, $\varepsilon_t \sim \text{i.i.d. } F(0, 1)$, onde supomos (por simplicidade) $\mu_t = 0$, ou que μ_t já tenha sido estimada através de algum modelo condicional para a média, por exemplo, algum modelo da classe ARFIMA(p, d, q). A média não condicional deste processo é $E[r_t] = E[E[r_t | I_{t-1}]] = 0$. A variância não condicional é $\text{var}(r_t) = E[r_t^2] = E[E[r_t^2 | I_{t-1}]] = E[\sigma_t^2] = \alpha_0 + \alpha_1 E[r_{t-1}^2]$. Sendo o processo estacionário de segunda ordem obtemos

$$\gamma_0 \equiv \sigma^2 \equiv \text{var}(r_t) = \frac{\alpha_0}{1 - \alpha_1},$$

e como a variância deve ser limitada, positiva, devemos ter $0 \leq \alpha_1 < 1$. Devemos também obter as covariâncias, $\text{cov}(r_t, r_{t+k}) = E[r_t r_{t+k}] = E[E[r_t r_{t+k} | I_{t+k-1}]] = E[r_t E[r_{t+k} | I_{t+k-1}]] = 0$. Isto é

$$\gamma_k \equiv \text{cov}(r_t, r_{t+k}) = 0 \quad \text{para } k \geq 1,$$

e portanto temos um processo de v.a.'s não correlacionadas, com média zero, e variância $\frac{\alpha_0}{1-\alpha_1}$ constante, *um ruído branco*. Finalmente, assumindo F como a distribuição Normal, obtemos o quarto momento. Sabendo que $E[r_t^4 | I_{t-1}] = E[\sigma_t^4 \varepsilon_t^4 | I_{t-1}] = 3E[\sigma_t^4 | I_{t-1}] = 3E[(\alpha_0 + \alpha_1 r_{t-1}^2)^2 | I_{t-1}]$, temos $E[r_t^4] = 3(\alpha_0^2 + 2\alpha_0\alpha_1 \text{var}(r_{t-1}) + \alpha_1^2 E[r_{t-1}^4])$, e como o processo é estacionário

$$E[r_t^4] = \frac{3\alpha_0^2(1 + \alpha_1)}{(1 - \alpha_1)(1 - 3\alpha_1^2)}.$$

Assim, para que o momento de quarta ordem seja finito e positivo devemos ter $(1 - 3\alpha_1^2) > 0$, ou seja, $0 \leq \alpha_1^2 < \frac{1}{3}$. O coeficiente de curtose do processo ARCH(1) é então

$$C = \frac{E[r_t^4]}{(\text{var}(r_t))^2} = \dots = 3 \frac{1 - \alpha_1^2}{1 - 3\alpha_1^2} > 3,$$

mesmo tendo inovações Normais.

Pode-se mostrar que um ARCH(1) corresponde a um AR(1) para os quadrados dos retornos. Para isto tome $r_t^2 - \sigma_t^2$ e equacione

$$r_t^2 - (\alpha_0 + \alpha_1 r_{t-1}^2) = \sigma_t^2 \varepsilon_t^2 - \sigma_t^2 = \sigma_t^2 (\varepsilon_t^2 - 1)$$

isto é

$$r_t^2 = \sigma_t^2 (\varepsilon_t^2 - 1) + (\alpha_0 + \alpha_1 r_{t-1}^2)$$

e obtemos o AR(1)

$$r_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \vartheta_t$$

onde $\vartheta_t = \sigma_t^2 (\varepsilon_t^2 - 1)$, ou ϑ_t é a v.a. $\sigma_t^2 (\chi_1^2 - 1)$. Portanto $\{\vartheta_t\}$ é uma sequência de v.a.'s não correlacionadas com média zero, mas com variância não constante. A f.a.c. de r_t^2 decai como a de um AR(1) com $\rho_k(r_t^2) = \alpha_1^k$.

Analogamente pode-se mostrar que um modelo ARCH(m) corresponde a um AR(m) para os quadrados dos retornos.

$$r_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \dots + \alpha_m r_{t-m}^2 + \vartheta_t,$$

onde $\{\vartheta_t\}$ é um ruído branco, o que irá justificar o teste abaixo e o uso da f.a.c.p. como ferramenta exploratória.

Para o ajuste de um modelo da família (G)ARCH(m) a uma série de retornos financeiros, podemos iniciar examinando a f.a.c.p. amostral da série $\{r_t^2\}$. Esta inspeção nos dará uma proposta inicial para a ordem m . Pode-se também efetuar o teste Multiplicadores de Lagrange (M.L.) proposto por Engle (1982). Este teste considera a regressão linear $r_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \dots + \alpha_m r_{t-m}^2 + e_t$, $t = 1, \dots, T$, ajusta o modelo por MQO, obtém a estatística F que tem assintoticamente a distribuição chiquadrado com m graus de liberdade, χ_m^2 , sendo $F = \frac{(SSR_0 - SSR_1)/m}{SSR_1/(T-2m-1)}$, onde SSR_0 é a soma de quadrados total, SSR_1 é a soma dos quadrados dos resíduos, $SSR_0 - SSR_1$ representa a variabilidade explicada pela regressão, e onde a hipótese nula é $H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_m = 0$. A hipótese nula é rejeitada para valores grandes da estatística teste e significa que “existe efeito ARCH”. Vale a pena lembrar que o retorno ao quadrado é um estimador viciado para a variância (super-estima) e não é eficiente, não sendo portanto uma boa proxy para avaliar a performance da volatilidade GARCH estimada, mas serve para especificar as dependências lineares de r_t^2 em $r_{t-1}^2, \dots, r_{t-m}^2$, o que justifica o uso da f.a.c.p.

A estimação de modelos de volatilidade é em geral feita pelo método da máxima verossimilhança, e assim é preciso especificar a densidade f das inovações. Supondo a média zero, a função de verossimilhança é

$$f(r_1, r_2, \dots, r_T | (\alpha_0, \alpha_1, \dots, \alpha_m)) = f(r_T | I_{T-1}) f(r_{T-1} | I_{T-2}) \dots f(r_{m+1} | I_m) g(r_1, r_2, \dots, r_m | (\alpha_0, \alpha_1, \dots, \alpha_m))$$

onde g é uma densidade conjunta. As estimativas de σ_t^2 são obtidas recursivamente. A densidade g tem em geral forma complicada e é usualmente retirada do processo de maximização, dando origem aos EMV condicionais. Por exemplo, supondo F Normal

$$f(r_{m+1}, r_{m+2}, \dots, r_T | (\alpha_0, \alpha_1, \dots, \alpha_m), (r_1, \dots, r_m)) = \prod_{t=m+1}^T \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left[\frac{-r_t^2}{2\sigma_t^2}\right].$$

Sob normalidade a log-verossimilhança é

$$l(r_{m+1}, r_{m+2}, \dots, r_T) = \sum_{t=m+1}^T -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(\sigma_t^2) - \frac{1}{2} \left(\frac{r_t}{\sigma_t}\right)^2.$$

Um ajuste clássico (EMV) pode produzir estimativas infladas para σ_t^2 . Um ajuste robusto deveria limitar as quantidades $\ln(\sigma_t^2)$ e $(\frac{r_t}{\sigma_t})^2$, podendo produzir estimativas mais realistas, veja Mendes, B. V. M. (2000a) e Muler, N. e Yohai, V. (2008).

Os *pseudo* estimadores de máxima verossimilhança são aqueles obtidos a partir da suposição de normalidade das inovações quando de fato elas seguiriam uma outra distribuição. A distribuição t_ν é uma opção melhor e está em geral implementada nos pacotes usuais. Neste caso a inovação padrão com média zero e variância um é obtida dividindo-se a v.a. t_ν por seu desvio padrão $\sqrt{\frac{\nu}{\nu-2}}$. Para ν previamente fixado, obtemos após alguns passos a log-verossimilhança

$$l(r_{m+1}, r_{m+2}, \dots, r_T) = - \sum_{t=m+1}^T \left[\frac{\nu+1}{2} \ln\left(1 + \frac{1}{(\nu-2)} \left(\frac{r_t}{\sigma_t}\right)^2\right) + \frac{1}{2} \ln(\sigma_t^2) \right],$$

onde quantidades similares as acima podem ser robustificadas.

Todos os passos sugeridos para o ajuste de um modelo ARFIMA devem ser aqui repetidos. Eles são, nesta ordem, identificação da ordem do modelo utilizando testes e as f.a.c. e f.a.c.p. amostrais, estimação do modelo escolhido, testes de significância das estimativas dos parâmetros, análise dos resíduos padronizados $\frac{r_t}{\sigma_t}$, ou de $\frac{r_t - \mu_t}{\sigma_t}$ se a média condicional for estimada conjuntamente, e uma possível reformulação do modelo. O critério de Akaike pode ser usado para decidir entre modelos. A verificação feita nos resíduos pode incluir um teste KS, um qq-plot, aplicação do teste de M.L. nos resíduos, aplicação do teste de Ljung-Box nos resíduos e resíduos ao quadrado e o exame da f.a.c. e f.a.c.p. dos mesmos. É sempre interessante ver o gráfico no tempo da volatilidade estimada, tentando identificar crises, e também examinar no tempo o gráfico dos resíduos padronizados, observando que os *clusters* de volatilidade desapareceram mas que os pontos extremos ainda estão lá. Isto nos motiva a usar modelos da TVE (Capítulo 2) para estimar com precisão as caudas da distribuição desses resíduos padronizados para, por exemplo, usar no cálculo de medidas de risco (Capítulo 7). Vale a pena também notar que uma estimação robusta dos modelos (G)ARCH (ou pelo menos o uso de uma distribuição com caudas pesadas) irá implicar em variâncias condicionais *não* infladas (ou menos infladas), o que pode destacar ainda mais os pontos atípicos.

As previsões da volatilidade são obtidas recursivamente como nos modelos AR(p). Seja a origem h das previsões. A previsão 2-passos-a-frente é estimada a partir da previsão 1-passo-a-frente $\widehat{\sigma}_h^2(1)$

$$\widehat{\sigma}_h^2(2) = \widehat{\alpha}_0 + \widehat{\alpha}_1 \widehat{\sigma}_h^2(1) + \widehat{\alpha}_2 r_h^2 + \dots + \widehat{\alpha}_m r_{h-m+2}^2,$$

e portanto a previsão l passos a frente é estimada como

$$\widehat{\sigma}_h^2(l) = \widehat{\alpha}_0 + \sum_{j=1}^m \widehat{\alpha}_j \widehat{\sigma}_h^2(l-j),$$

onde $\widehat{\sigma}_h^2(l-j) = r_{h+l-j}^2$ se $l-j \leq 0$.

Agora um exemplo. Considere a série de retornos do índice brasileiro. Primeiro efetuamos o teste M.L. para testar a existência de efeito ARCH conforme sugerido em Engle (1982). A hipótese nula de “Não Existe Efeito ARCH” foi rejeitada com p-valor zero, e o melhor ajuste ARCH foi

```
Mean Equation: br ~ + arma(1, 1) - 1
Conditional Distribution: t with parameter 9
```

```
-----
Estimated Coefficients:
```

```
-----
              Value Std.Error   t value Pr(>|t|)
AR(1) -0.96056    0.02559  -37.5331 0.000000
MA(1)  0.97742    0.01931   50.6050 0.000000
A      1.94386    0.19190   10.1298 0.000000
ARCH(1) -0.01447   0.02281   -0.6346 0.525794
ARCH(2)  0.08061   0.03484    2.3137 0.020835
ARCH(3)  0.12693   0.04126    3.0765 0.002136
ARCH(4)  0.11180   0.04293    2.6043 0.009306
ARCH(5)  0.06747   0.03477    1.9403 0.052546
-----
```

```
AIC(8) = 5284.077
```

Este certamente não é o melhor modelo para a volatilidade desta série. A especificação de uma distribuição condicional t combinada com extensões deste modelo devem melhorar o ajuste. Notemos ainda que as estimativas dos parâmetros ARMA são no mínimo suspeitas.

Alguns autores apontam algumas deficiências do modelo ARCH: (i) choques positivos e negativos têm o mesmo efeito na volatilidade; (ii) algumas restrições complicadas para os α_i 's; (iii) fornecem apenas uma fórmula para descrever o mecanismo gerador dos clusters de volatilidade, não dando nenhuma "pista" ou indicação das causas de tal comportamento; (iv) grande sensibilidade à observações atípicas, isto é, grandes choques que poderiam ser "outliers" causam super-previsão de σ_t . Muitas das extensões que se seguiram foram concebidas com o intuito de corrigir essas falhas.

Modelo GARCH(m, s)

O modelo *Generalized autoregressive conditionally heteroskedastic* GARCH, proposto por Bollerslev, T. (1986), Bollerslev, T. (1987), expressa de maneira mais parcimoniosa a estrutura de dependência da variância condicional, já que é fato comprovado empiricamente que um ARCH(m) não ajusta bem séries de retornos financeiros a menos que m seja grande. Nesse modelo, a variância condicional, além de depender dos quadrados dos retornos passados como no modelo ARCH, depende também das próprias variâncias condicionais passadas.

Se $\{r_t\}$ segue um modelo GARCH(m, s), então $r_t = \sqrt{\sigma_t^2} \varepsilon_t$ e

$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \cdots + \alpha_m r_{t-m}^2 + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-2}^2 + \cdots + \beta_s \sigma_{t-s}^2, \quad (5.5)$$

com $\varepsilon_t \sim \text{i.i.d.} F(0, 1)$, $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, \dots, m$, $\beta_i \geq 0$, $i = 1, \dots, s$, e onde a *persistência* $\sum_{j=1}^q (\alpha_j + \beta_j)$ deve ser menor que 1 para que a variância não condicional seja finita, $q = \max(m, s)$. Na prática, novamente, supomos F sendo a distribuição Normal ou t_ν .

Da mesma forma como fizemos nos modelos ARCH, podemos escrever o GARCH como um ARMA(q, s) nos quadrados dos retornos, r_t^2 . Para isto, seja $\eta_t = r_t^2 - \sigma_t^2$ e considere a equação (5.5). Substitua em (5.5) os σ_{t-j}^2 por $r_{t-j}^2 - \eta_{t-j}$, $j = 0, 1, \dots, m$, obtendo

$$r_t^2 = \alpha_0 + \sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i) r_{t-i}^2 + \eta_t - \sum_{j=1}^s \beta_j \eta_{t-j}.$$

Como $E[\eta_j] = 0$ e $\text{cov}(\eta_j, \eta_{j+k}) = 0$ para $k \geq 1$ (um $\text{RB}(0, \sigma_\eta^2)$), podemos ver a última expressão como um modelo ARMA para a série $\{r_t^2\}$. Como a expressão da variância não condicional de um processo ARMA é conhecida (estamos supondo retornos com média zero), temos $E[r_t^2] = \frac{\alpha_0}{1 - \sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i)}$, cujo denominador deve ser sempre positivo.

A estimação dos modelos GARCH(m, s) segue os mesmos passos vistos para o ARCH(m), devendo começar com a determinação das ordens m e s . Para isto podemos examinar a f.a.c.p. amostral (da série ao quadrado) e também aplicar o teste de M.L. para detectar efeito ARCH. É indiferente se usar os retornos ou os retornos filtrados (resíduos) por um modelo ARMA(p, q), já que o interesse é no segundo momento. Podemos também aplicar o teste de Ljung-Box na série ao quadrado.

Rejeição de ambas as hipóteses nulas acima, nos leva ao segundo passo que é a estimação do modelo. Uma abordagem usual é superparametrizar, e ir aos poucos retirando os coeficientes não significativos do modelo, e utilizando o critério AIC para decidir entre os modelos (desde que os modelos competidores contenham todas as estimativas dos parâmetros significativas). Como no caso ARCH, em geral o método de estimação utilizado é o de máxima verossimilhança. No caso dos erros seguirem a distribuição t_ν , os graus de liberdade devem também ser estimados. A escolha de uma distribuição F com caudas pesadas sempre melhora o ajuste e reduz o número de parâmetros.

Bollerslev, T. e Wooldridge, J. M. (1992) mostraram que os *pseudo* EMV (EMV assumindo normalidade mesmo quando esta pode não ser a verdadeira distribuição dos erros) dos parâmetros do modelo são consistentes. Contudo os estimadores da variância desses estimadores não são consistentes se a suposição de normalidade é violada. Veja na referência dada uma metodologia para obter erros padrões consistentes.

Finalmente, devemos examinar os resíduos do melhor modelo ajustado. Os choques padronizados $\{\hat{\varepsilon}_t\}$ devem ser i.i.d. $F(0, 1)$. Podemos usar os testes de Ljung-Box, de assimetria, de curtose, e também examinar os q-q-plots, e as f.a.c. e f.a.c.p. Se alguma suposição não se verificar, devemos voltar ao primeiro passo e procurar por um modelo melhor. Somente após todas essas verificações podemos seguir com as estimativas das previsões do modelo. Além disto, é sempre interessante observar o gráfico das volatilidades *in-sample*.

Modelo GARCH(1, 1)

O modelo mais simples nesta família é o GARCH(1,1). A beleza dessa especificação reside no fato

de fornecer um modelo parcimonioso que, em geral, explica muito bem a dinâmica da volatilidade de retornos financeiros. Por exemplo, para a série de retornos do índice brasileiro, para a qual achamos a ordem $m = 5$ com AIC= 5284.077, mantendo-se fixo todo o restante da especificação, temos que o ajuste GARCH(1,1) resultou em um AIC menor. Veja que a parte ARMA não é mais significativa e que a estimativa do α_1 é agora estatisticamente significativa.

```
Mean Equation: br ~ + arma(1, 1) - 1
Conditional Variance Equation: ~ garch(1, 1)
Conditional Distribution: t with parameter 9
```

Estimated Coefficients:

	Value	Std.Error	t value	Pr(> t)
AR(1)	-0.5850	0.61458	-0.9519	0.341297
MA(1)	0.6080	0.60042	1.0126	0.311449
A	0.0799	0.04399	1.8163	0.069540
ARCH(1)	0.0512	0.01583	3.2334	0.001253
GARCH(1)	0.9246	0.02570	35.9692	0.000000

AIC(5) = 5273.792

Para séries de retornos financeiros em geral basta um modelo GARCH(1,1) para capturar a dinâmica da volatilidade da série. Supondo normalidade para as inovações, temos que para o GARCH(1,1) a curtose C é

$$C(r_t) = \frac{E[r_t^4]}{(E[r_t^2])^2} = \frac{3(1 - (\alpha_1 + \beta_1)^2)}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2} > 3.$$

Um processo GARCH(1,1) tem portanto caudas mais pesadas que a Normal mesmo quando as inovações são Normais.

Podemos calcular a expressão da *forward persistence*, ou simplesmente persistência, fórmula (5.2) para o GARCH(1,1). Primeiro notemos que sendo $\gamma_0 = \frac{\alpha_0}{(1-\alpha_1-\beta_1)}$, temos $\alpha_0 = (1-\alpha_1-\beta_1)\gamma_0$. Temos então que

$$\sigma_t^2 - \gamma_0 = \alpha_1(r_{t-1}^2 - \gamma_0) + \beta_1(\sigma_{t-1}^2 - \gamma_0)$$

Sendo $\varepsilon_t \equiv i.i.d.(0, 1)$ temos que $E[r_t^2 | \mathcal{I}_{t-1}] = E[\sigma_t^2 \varepsilon_t^2 | \mathcal{I}_{t-1}] = \sigma_t^2$. Pode-se facilmente mostrar que

$$E[r_{t+k}^2 | \mathcal{I}_t] = \sigma_{t+k}^2.$$

Assim, para $k \geq 2$

$$E[(\sigma_{t+k}^2 - \gamma_0) | \mathcal{I}_t] = (\alpha_1 + \beta_1)E[(\sigma_{t+k-1}^2 - \gamma_0) | \mathcal{I}_t].$$

Após algumas iterações chegamos a

$$\sigma_{t+k|t}^2 = \gamma_0 + (\alpha_1 + \beta_1)^{k-1}(\alpha_0 + \alpha_1 r_t^2 + \beta_1 \sigma_t^2 - \gamma_0)$$

e assim

$$\theta_{t+k|t} = \alpha_1(\alpha_1 + \beta_1)^{k-1}.$$

Obtemos também a persistência acumulada

$$\phi_{t+k|t} = \frac{1}{k}(\theta_{t+k|t} + \theta_{t+k-1|t} + \cdots + \theta_{t+1|t})$$

que para o GARCH(1,1) é

$$\phi_{t+k|t} = \frac{1}{k}\alpha_1 \frac{1 - (\alpha_1 + \beta_1)^k}{1 - \alpha_1 - \beta_1}.$$

A meia vida da volatilidade

$$k = \frac{\log((\alpha_1 + \beta_1)/2)}{\log(\alpha_1 + \beta_1)},$$

já que $|\gamma_0 + (\alpha_1 + \beta_1)^{k-1}(\sigma_{t+1}^2 - \gamma_0) - \gamma_0| = 0.5 |\sigma_{t+1}^2 - \gamma_0|$ o que implica em $(\alpha_1 + \beta_1)^{k-1} = 0.5(\alpha_1 + \beta_1)$.

A previsão 1 passo a frente do horizonte h é dada por

$$\widehat{\sigma_{h+1}^2} = \widehat{\sigma_h^2}(1) = \widehat{\alpha_0} + \widehat{\alpha_1}r_h^2 + \widehat{\beta_1}\sigma_h^2$$

onde σ_h^2 é obtida *in-sample*. A previsão 2 passos a frente do horizonte h

$$\widehat{\sigma_h^2}(2) = \widehat{\alpha_0} + (\widehat{\alpha_1} + \widehat{\beta_1})\widehat{\sigma_h^2}(1).$$

Para chegar na expressão acima basta ver que $\sigma_{t+1}^2 = \alpha_0 + \alpha_1 r_t^2 + \beta_1 \sigma_t^2$, e como $r_t^2 = \sigma_t^2 \varepsilon_t^2$, somando e subtraindo $\alpha_1 \sigma_t^2$ temos que $\sigma_{t+1}^2 = \alpha_0 + \sigma_t^2(\alpha_1 + \beta_1) + \alpha_1 \sigma_t^2(\varepsilon_t^2 - 1)$. Quando $t = h+1$, esta expressão fica

$$\sigma_{h+2}^2 = \alpha_0 + \sigma_{h+1}^2(\alpha_1 + \beta_1) + \alpha_1 \sigma_{h+1}^2(\varepsilon_{h+1}^2 - 1),$$

mas como $E[\varepsilon_{h+1}^2 - 1 | \mathcal{I}_h] = 0$, chegamos na previsão dois passos a frente $\widehat{\sigma_h^2}(2)$ dada acima.

Generalizando, para $l > 1$ temos $\widehat{\sigma_h^2}(l) = \alpha_0 + \alpha_1 \widehat{\sigma_h^2}(l-1) + \beta_1 \widehat{\sigma_h^2}(l-1)$, e após algumas contas chegamos a

$$\widehat{\sigma_h^2}(l) = \widehat{\alpha_0} + (\widehat{\alpha_1} + \widehat{\beta_1})\widehat{\sigma_h^2}(l-1) \quad l > 1.$$

Por substituições sucessivas nesta última expressão podemos escrever as previsões como

$$\widehat{\sigma_h^2}(l) = \frac{\alpha_0(1 - (\alpha_1 + \beta_1)^{l-1})}{1 - \alpha_1 - \beta_1} + (\alpha_1 + \beta_1)^{l-1} \widehat{\sigma_h^2}(1).$$

Quando l aumenta, a quantidade $(\alpha_1 + \beta_1)^{l-1}$ tende a zero e, portanto,

$$\widehat{\sigma_h^2}(l) \rightarrow \frac{\alpha_0}{1 - \alpha_1 - \beta_1}, \text{ quando } l \rightarrow \infty,$$

desde que $\alpha_1 + \beta_1 < 1$. Assim, se a $\text{var}(r_t)$ existir, temos que as previsões das variâncias condicionais l -passos a frente converge para a variância não condicional de r_t , quando $l \rightarrow \infty$, fato conhecido como

a reversão para a média de r_t^2 , ou para a variância não condicional de r_t . Portanto, para horizontes muito distantes pode ser que não haja vantagem em se fazer previsões condicionais.

Não há como de fato verificar a qualidade das previsões já que a volatilidade não é observável. Podemos esperar que o melhor modelo *in-sample* forneça também as melhores estimativas da volatilidade. Fora da amostra, pode-se comparar com algumas proxies, e/ou comparar a performance de medidas de risco baseadas em diferentes estimativas da volatilidade.

Um critério bastante utilizado consiste em procurar pelo mínimo da função de perda $\sum_{t=1}^N (\widehat{\sigma}_t^2 - r_t^2)$. Mas, conforme já observamos, r_t^2 é um estimador viciado e não eficiente da volatilidade. De fato, alguns artigos concluíram sobre uma performance ruim das previsões GARCH, mas isto deveu-se ao uso da proxy r_t^2 . Na verdade, Andersen, T. G. e Bollerslev, T. (1998a) usando como proxy uma medida de volatilidade realizada baseada em dados de alta frequência mostrou que as previsões GARCH são na verdade muito boas. Medidas de volatilidade realizada são estimadores consistentes, mas sua obtenção requer muitos cuidados com os dados de alta frequência que devem ser filtrados para retirada de efeitos de micro-estrutura, ciclos, memória longa, etc. A inclusão de variáveis exógenas na equação da variância condicional pode também melhorar as previsões.

Existe hoje uma enorme variedade de extensões dos modelos GARCH, motivadas pelos fatos estilizados observados. Destacamos aqui os modelos GARCH-M, EGARCH, FIGARCH, FIEGARCH, TGARCH, SW-ARCH, LM-ARCH, o *hyperbolic* GARCH, ou HYGARCH de Davidson (2004), entre vários outros. Conforme vimos, os modelos ARCH e GARCH tratam simetricamente os retornos, no sentido que retornos positivos e negativos não são distinguíveis, já que a variância condicional σ_t^2 é função dos retornos ao quadrado. A prática, no entanto, nos mostra que a volatilidade é maior após um retorno negativo de magnitude grande. Esta característica é modelada pelos processos EGARCH e TGARCH.

Modelo EGARCH(m, s)

O modelo exponencial GARCH, foi proposto por Nelson, D. B. (1991). É um modelo interessante porque evita alguns dos problemas de otimização encontrados pelos modelos GARCH, especificamente, a possibilidade de obter estimativas negativas para a variância (e assim a necessidade de se impor restrições para os parâmetros no algoritmo), e o tratamento simétrico dado aos choques positivos e negativos.

O modelo EGARCH(m, s) pode ser escrito como

$$\log(\sigma_t^2) = \alpha_0 + \sum_{i=1}^m [\alpha_i (|\frac{r_{t-i}}{\sigma_{t-i}}| - E|\frac{r_{t-i}}{\sigma_{t-i}}|) + \gamma_i \frac{r_{t-i}}{\sigma_{t-i}}] + \sum_{j=1}^s \beta_j \log(\sigma_{t-j}^2) \quad (5.6)$$

onde γ_i captura o efeito de assimetria da informação.

Se r_t segue um modelo EGARCH(1, 1) então $r_t = \sigma_t \varepsilon_t$ e

$$\log(\sigma_t^2) = \alpha_0 + \alpha_1 g(\varepsilon_{t-1}) + \beta_1 \log(\sigma_{t-1}^2)$$

onde $\varepsilon_t \sim \text{i.i.d. F}(0, 1)$, α_0 é uma constante (podendo ser negativa), e $g(\cdot)$ é a curva de impacto de informação, onde $\gamma \in \mathbb{R}$, e $|\varepsilon_t| - E[|\varepsilon_t|]$ é uma sequência de v.a.'s i.i.d. com média zero. A função

$g(\varepsilon_t)$ esclarece o tratamento assimétrico dado aos retornos. Assim, o modelo dá respostas diferentes aos retornos positivos e negativos, fato conhecido como efeito *leverage*.

Notemos que se $\varepsilon_t \sim N(0, 1)$, então $E[|\varepsilon_t|] = \sqrt{\frac{2}{\pi}}$. Se $\varepsilon_t \sim \frac{t_\nu}{\sqrt{\nu/(\nu-2)}}$, v.a. com média zero e variância um, já que t_ν representa uma v.a. t-student com ν graus de liberdade e variância $\frac{\nu}{\nu-2}$, então $E[|\varepsilon_t|] = \frac{2\sqrt{\nu-2}\Gamma(\frac{\nu+1}{2})}{(\nu-1)\Gamma(\nu/2)\sqrt{\pi}}$.

Um modelo EGARCH(m, s) (com esperança não condicional zero) pode ter sua volatilidade reescrita como

$$\log(\sigma_t^2) = \alpha_0 + \frac{1 + \beta_1 B + \dots + \beta_s B^s}{1 - \alpha_1 B - \dots - \alpha_m B^m} g(\varepsilon_{t-1}),$$

onde os polinômios $1 + \beta_1 B + \dots + \beta_s B^s$ e $1 - \alpha_1 B - \dots - \alpha_m B^m$ têm raízes fora do círculo unitário. Note que esta é uma representação ARMA para a evolução da variância condicional de r_t . Temos ainda que a esperança não condicional $E[\log(\sigma_t^2)] = \alpha_0$. No SPlus o modelo EGARCH(1,1) é escrito como

$$\log(\sigma_t^2) = w + \beta \log(\sigma_{t-1}^2) + \alpha \left| \frac{r_{t-1}}{\sigma_{t-1}} \right| + \gamma \frac{r_{t-1}}{\sigma_{t-1}}.$$

Note γ pode ser zero. Quando $\gamma \neq 0$ temos o efeito *leverage*.

Modelo TGARCH(m, s)(Threshold ARCH)

Tem como caso particular o ARCH. A volatilidade agora é dada por

$$\sigma_{t,\gamma}^2 = \alpha_0 + \alpha_1 g_\gamma(\varepsilon_{t-1}) + \beta_1 \sigma_{t-1,\gamma}^2$$

onde

$$g_\gamma(\varepsilon_t) = \theta \mathbf{1}_{[\varepsilon_t > 0]} |\varepsilon_t|^\gamma + (1 - \theta) \mathbf{1}_{[\varepsilon_t \leq 0]} |\varepsilon_t|^\gamma.$$

Se $\gamma = 1$ temos o modelo de Zakoian (1994). Se $\gamma = 2$, temos o modelo GJR (Glosten, L. R., Jagannathan, R. e Runkle, D. E., 1993). Nos *softwares* usados possuem a formulação

$$\sigma_t^2 = w + \alpha r_{t-1}^2 + \gamma r_{t-1}^2 d_{t-1} + \beta \sigma_{t-1}^2$$

onde $d_t = 1$ se $r_t < 0$, e zero caso contrário, para captar o efeito das *bad news*. Se $\gamma \neq 0$ existe assimetria da informação, sendo que se γ estatisticamente maior que zero, significa que as notícias ruins tem efeito maior que as boas.

Modelo IGARCH(m, s) (GARCH Integrado)

Assim como o polinômio associado ao processo AR(p) podia ter uma raiz unitária (RU), a representação ARMA dos modelos GARCH também pode apresentar RU no polinômio da parte AR. Neste caso temos GARCH com RU, ou IGARCH. Seja $r_t^2 = \alpha_0 + (\alpha_1 + \beta_1)r_{t-1}^2 + \eta_t - \beta_1\eta_{t-1}$. Sendo $\alpha_1 + \beta_1 = 1$ temos que o impacto ou efeito de choques passados no r_t^2 é persistente.

Um IGARCH(1,1) pode ser escrito como

$$r_t = \sigma_t \varepsilon_t \quad \sigma_t^2 = \alpha_0 + \beta_1 \sigma_{t-1}^2 + (1 - \beta_1) r_{t-1}^2$$

onde $\beta_1 > 0$, $\beta_1 < 1$, e $\{\varepsilon_t\}$ é uma sequência de v.a.'s i.i.d.F(0,1). Assim como nos modelos ARIMA, o impacto de choques passados ao quadrado é persistente. A variância não condicional de r_t não está definida. Pode-se mostrar que a previsão l passos a frente a partir do horizonte h é

$$\sigma_h^2(l) = \sigma_h^2(1) + (l-1)\alpha_0 \quad l \geq 1$$

e portanto o efeito de $\sigma_h^2(1)$ nas volatilidades futuras é persistente, isto é, afeta todas as previsões a serem feitas. Note que as previsões formam uma reta com inclinação α_0 . O caso particular onde $\mu = 0$ é o modelo adotado em *Riskmetrics* para o cálculo do VaR.

Modelo GARCH-M(m, s) (GARCH-in-Mean)

Este modelo se adequa a situações onde o valor do retorno de um ativo depende de sua volatilidade. O GARCH-M(1, 1) é especificado como

$$r_t = \mu + c\sigma_t^2 + a_t, \quad a_t = \sigma_t \varepsilon_t$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

onde μ e c são constantes. O parâmetro c é chamado *prêmio pelo risco*. Notemos que c positivo indica que o retorno é positivamente relacionado com sua variância condicional passada, sugerindo que “uma variância grande ontem implicaria em um retorno maior hoje”. Na especificação da média condicional, pode-se usar, alternativamente, σ_t ou $\log(\sigma_t^2)$.

É interessante ver que esta formulação diz que existe correlação serial em r_t e que estes ρ_k 's de r_t seriam induzidos pela existência de autocorrelação serial no processo $\{\sigma_t^2\}$. Em outras palavras, a existência de correlação serial em r_t seria explicada pela existência do “prêmio pelo risco”. É uma modelagem interessante que se aplica a outras extensões dos processos GARCH.

Abaixo mostramos o excelente ajuste para os retornos do índice brasileiro. Observe que o melhor modelo para a média agora é um ARMA(0, 0).

Mean Equation: `br ~ var.in.mean + arma(0, 0) - 1`

Conditional Variance Equation: `~ garch(1, 1)`

Conditional Distribution: `t with parameter 13`

Estimated Coefficients:

	Value	Std.Error	t value	Pr(> t)
ARCH-IN-MEAN	0.05096	0.01532	3.326	0.0009046
A	0.07961	0.03918	2.032	0.0423349
ARCH(1)	0.04992	0.01420	3.515	0.0004541
GARCH(1)	0.92332	0.02396	38.537	0.0000000

AIC(4) = 5256.884

Notemos ainda que a aplicação de modelos GARCH a séries muito longas pode resultar em uma medida de persistência muito alta (quase um I-GARCH) devido (talvez) à presença de saltos (mudanças de regime) determinísticas da variância não condicional que não foram devidamente modeladas (ou contempladas pela modelagem). Alguns autores trabalharam neste tópico tentando desenvolver meios de identificar o momento de tais mudanças, por exemplo, Lamoureux, C. G. e Lastrapes, W. D. (1990).

Novamente ressaltamos que existem inúmeras outras variações dos modelos GARCH na literatura que não serão tratadas aqui, por exemplo os *Component* GARCH, veja Engle, R. F. e Lee, G. G. J. (1999). Uma extensão importante é aquela que considera a possibilidade de existência de memória longa na volatilidade, os modelos FIGARCH.

Processos FIGARCH(m, d, s)

Toda a teoria desenvolvida para os processos ARFIMA foi naturalmente estendida na direção dos modelos de volatilidade. Os modelos *Fractionally Integrated Generalized Autoregressive Conditionally Heteroskedastic* (FIGARCH) foram introduzidos por Baillie, R. T., Bollerslev, T. e Mikkelsen, H. O. (1996), e Bollerslev, T. e Mikkelsen, H. O. (1996), motivados pelo fato de que a função de autocorrelação dos quadrados dos retornos, ou do seu valor absoluto, decai lentamente mesmo quando a série não possui correlação serial.

Considere a formulação dos modelos GARCH, $r_t = \sigma_t \varepsilon_t$, onde ε_t é uma sequência i.i.d. com média zero e variância um, e seja \mathcal{F}_{t-1} a sigma-álgebra de eventos gerada pelas informações passadas. Para um processo FIGARCH a variância de $r_t | \mathcal{F}_{t-1}$ é dada por

$$\sigma_t^2 = \omega (1 - \beta(B))^{-1} + \lambda(B) r_t^2,$$

onde

$$\lambda(B) = \sum_{k=0}^{\infty} \lambda_k B^k = 1 - (1 - \beta(B))^{-1} \phi(B) (1 - B)^d, \quad (5.7)$$

onde $\phi(B) = 1 - \alpha(B) - \beta(B)$, onde $\omega > 0$ é uma constante real, $\alpha(B) = \sum_{i=1}^m \alpha_i B^i$ e $\beta(B) = \sum_{j=1}^s \beta_j B^j$. A expansão binomial em B é dada por

$$(1 - B)^d = 1 + \sum_{k=1}^{\infty} \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(-d)} B^k = 1 - d \sum_{k=1}^{\infty} \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(1-d)} B^k.$$

Em um processo FIGARCH(m, d, s) os parâmetros devem satisfazer restrições bastante complicadas para garantir que a variância seja sempre positiva. Por exemplo, para um FIGARCH($1, d, 1$), além da condição $\omega > 0$, deve-se ter $\beta_1 - d \leq \phi_1 \leq \frac{2-d}{3}$; e $d(\phi_1 - \frac{1-d}{2}) \leq \beta_1(d + \alpha_1)$, onde $\phi_1 = \alpha_1 + \beta_1$. É interessante notar que o FIGARCH($1, d, 1$) equivale a um GARCH($1, 1$), quando $d = 0$, e a um IGARCH($1, 1$) quando $d = 1$.

Aplicamos os testes para detectar memória longa nos quadrados dos retornos dos índices sob análise. Ambos os testes *RS* modificado e o GPH deram altamente significativos para o índice brasileiro ($d = 0.454$). Veja o decaimento da f.a.c. amostral na Figura 5.2.

Alguns autores têm apontado algumas deficiências do modelo FIGARCH. Por exemplo, Davidson, J. (2004) mostrou que a persistência nos choques na volatilidade de um processo FIGARCH decresce

quando o parâmetro de memória longa aumenta. Ruiz e Pérez (2003) mostraram que o modelo proposto por Hwang (2001) tem problemas de identificabilidade. A estimação do modelo para ordens maiores também apresenta problemas de convergência. Existem várias propostas de variações de modelos tipo ARCH que representam simultaneamente o efeito *leverage* e a memória longa. Por exemplo, Hwang (2001) propôs estender o modelo de memória longa FIGARCH de Baillie et al. (1996) para também representar o efeito *leverage*. O modelo FIEGARCH (seriam os processos EGARCH com memória longa, que considera o log da volatilidade) é computacionalmente mais estável e produz melhores estimativas.

5.3 Modelos de volatilidade estocástica com memória longa

Existem duas grandes famílias de modelos econométricos para representar a evolução dinâmica da volatilidade. Os modelos da família ARCH se caracterizam principalmente por especificar a volatilidade como uma função de potências dos valores absolutos dos retornos passados, e portanto seria observada um passo a frente. Por outro lado, os modelos de volatilidade estocástica (SV) especificam a volatilidade como uma variável latente não diretamente observável. Ambas as classes foram estendidas para incluir a possibilidade de memória longa na volatilidade.

No contexto de modelos tipo SV, Harvey (1998) e Breidt et al. (1998) independentemente propuseram o Modelo de Volatilidade Estocástica com Memória Longa (LMSV) no qual a log-volatilidade é modelada como um processo ARFIMA.

Seja $\{Y_t\}_{t=1}^n$ tal que

$$Y_t = g(X_t)\sigma_\varepsilon\varepsilon_t, \quad (5.8)$$

onde X_t é uma série temporal com média zero e memória longa, $g(\cdot)$ é uma função contínua e ε_t é uma série i.i.d. com média zero e variância um. Já que $\text{var}(Y_t|X_t) = g(X_t)^2\sigma_\varepsilon^2$, para certas funções $g(\cdot)$ o modelo (5.8) pode ser descrito como um processo de volatilidade estocástica com memória longa (Robinson, 1999). Esta grande classe de modelos de volatilidade inclui o modelo não linear média móvel de memória longa de Robinson e Zaffaroni (1998) e Zaffaroni (1999), e o processo FISV introduzido por Breidt et al. (1998).

Em um processo FISV($p, d, q, \sigma_\varepsilon$) a função $g(\cdot)$ em (5.8) é dada por

$$g(X_t) = \exp\left(\frac{X_t}{2}\right), \quad (5.9)$$

onde $\{X_t\}_{t \in \mathbb{Z}}$ é um processo ARFIMA(p, d, q) dado por (6.27), e ε_t e ϵ_t são i.i.d. Normais padrão mutuamente independente. Observe que $\text{var}(Y_t|X_t) = \exp(X_t)\sigma_\varepsilon^2$. Em particular, tomando o quadrado dos dois lados da equação (5.8) e tomando logaritmo,

$$\ln(Y_t^2) = \mu_\xi + X_t + \xi_t, \quad (5.10)$$

onde $\mu_\xi = \ln(\sigma_\varepsilon^2) + E[\ln(\varepsilon_t^2)]$, e $\xi_t = \ln(\varepsilon_t^2) - E[\ln(\varepsilon_t^2)]$. Então, $\ln(Y_t^2)$ é a soma de um processo ARFIMA Gaussiano e de um ruído não Normal com média zero. Consequentemente, a função de autocovariância do processo $\ln(Y_t^2)$, quando $d \in (-0.5, 0.5)$, é tal que

$$\gamma_{\ln(Y_t^2)}(k) \sim k^{2d-1}, \quad (5.11)$$

onde $k \rightarrow \infty$, enquanto sua função de densidade espectral tem a propriedade que

$$f_{\ln(Y_t^2)}(\lambda) \sim \lambda^{-2d}, \quad (5.12)$$

quando a frequência $\lambda \rightarrow 0$. Para $d \in (0.0, 0.5)$, a função de densidade espectral em (5.12) é ilimitada quando $\lambda \rightarrow 0$. Isto dá a base para a aplicação dos tradicionais procedimentos de estimação baseados no log-periodograma.

5.4 Ilustração

Daremos agora um exemplo analisando os conceitos e fatos discutidos sobre a volatilidade utilizando o índice do Brasil. A Figura 5.1 mostra os preços diários de fechamento e log-retornos percentuais do índice do Brasil de 03/01/2006 a 13/10/2011. Conforme discutido no Capítulo 1, a figura ilustra vários fatos estilizados sobre a evolução dos preços e dos retornos.

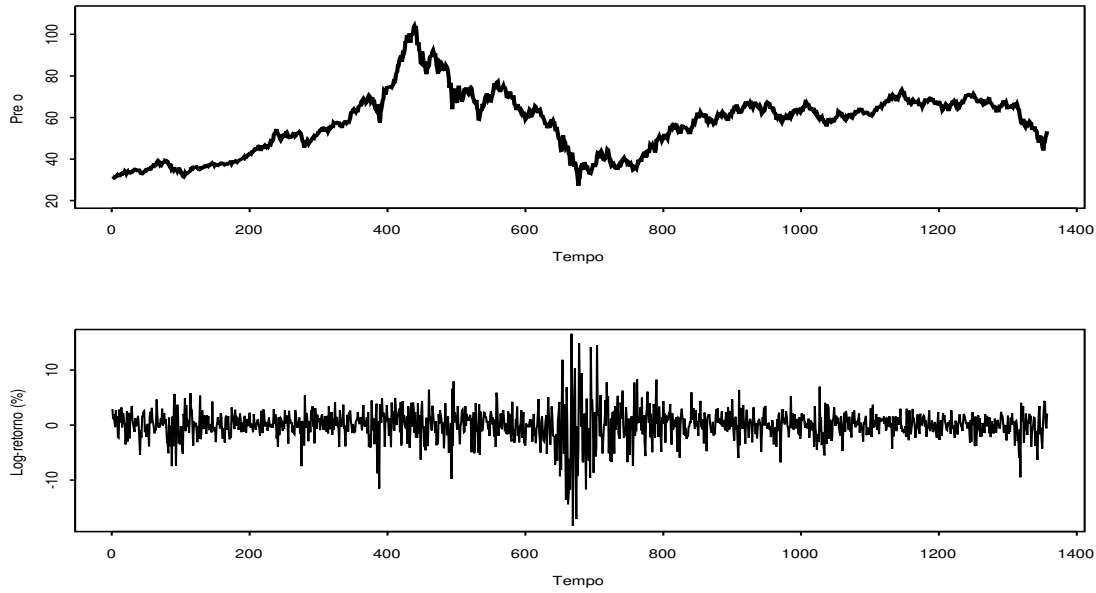


Figura 5.1: *Preços diários de fechamento e log-retornos percentuais do índice Brasil de 03/01/2006 a 13/10/2011.*

Apresentamos na Tabela 5.1 algumas estatísticas caracterizando os três primeiros momentos da distribuição dos log-retornos.

Como a tabela mostra, o índice brasileiro apresenta um log-retorno médio positivo e pequeno de aproximadamente 1/25 % por dia, mas a mediana é quase 5 vezes maior. A variância diária é 7.8140, o que implica numa volatilidade (desvio padrão) anualizada média de 44.37%. Observamos que a

Tabela 5.1: Estatísticas básicas dos log-retornos do índice Brasil.

Média amostral	0.0417
Mediana amostral	0.1873
Mínimo	-18.3232
Máximo	16.6188
Variância amostral	7.8140
Assimetria amostral	-0.4478
Curtose amostral	9.8233
T (tamanho da amostra)	1357

fórmula utilizada foi $volatilidade\ anualizada = \sqrt{252} * \sqrt{variância\ diaria}$. O coeficiente de assimetria indica que a distribuição dos retornos é bem assimétrica para a esquerda, como também indicam os valores do mínimo e máximo da amostra. A curtose, que mede o peso das caudas da distribuição é muito maior que o valor de referência 3 da distribuição Normal. Tudo isto indica que a distribuição dos log-retornos do índice não é a Normal, fato comprovado pelos testes de Jarque-Bera, Shapiro-Wilk e KS, todos apresentando p-valor zero.

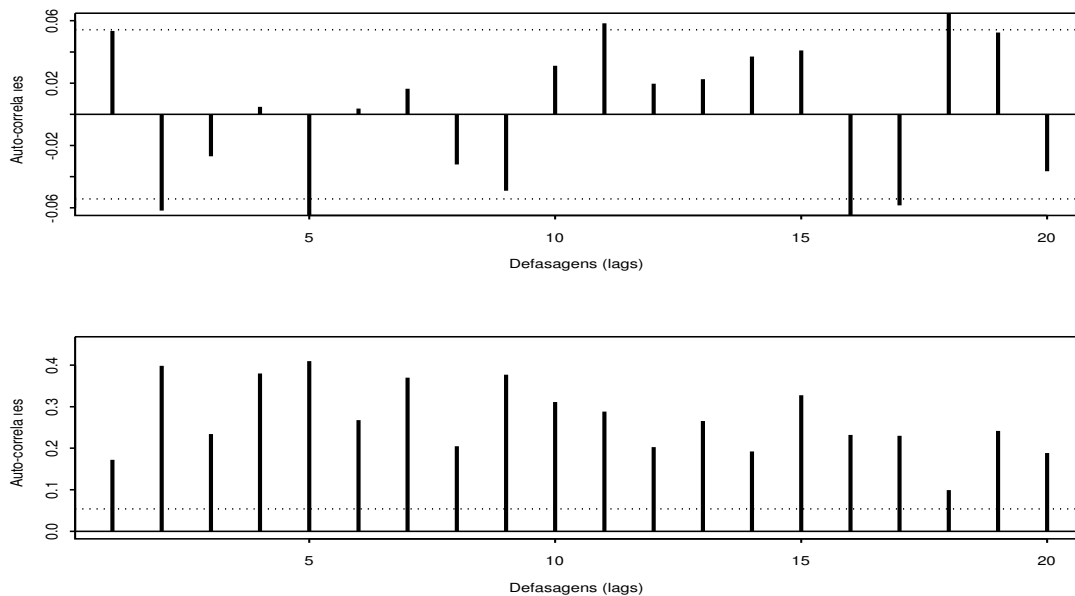


Figura 5.2: A f.a.c. amostral dos log-retornos percentuais e seus quadrados para o índice Brasil de 03/01/2006 a 13/10/2011.

A Figura 5.2 mostra a f.a.c. amostral dos log-retornos percentuais do índice brasileiro e dos seus

Tabela 5.2: *Estimativas (erros padrões) dos modelos GARCH para os log-retornos do índice Brasil. Todas estimativas significantes a 5%.*

Modelo	ϕ_0	α_0	α_1	β_1	LEV	$d(\text{ML})$	Distr.Cond.	AIC
M1:ARMA(0,0)-GARCH(1,1)	0.1235 (0.057)	0.1368 (0.036)	0.0999 (0.012)	0.8798 (0.015)	— —	— —	Normal —	6113.6 —
M2:ARMA(0,0)-GARCH(1,1)	0.1428 (0.054)	0.1103 (0.041)	0.0916 (0.016)	0.8919 (0.018)	— —	— —	t_5 —	6074.5 —
M3:ARMA(0,0)-FIGARCH(1,1)	0.1252 (0.056)	0.2008 (0.076)	0.0011 (0.070)	0.4676 (0.111)	— —	0.4676 (0.067)	Normal —	6108.8 —

quadrados. Os gráficos indicam que deve haver uma dependência fraca na média condicional, mas uma autocorrelação forte para várias defasagens na volatilidade dos retornos. As figuras também sugerem presença de memória longa na média e na volatilidade.

A seguir ajustamos um modelo ARFIMA(p, q)-GARCH(m, s) à esta série de retornos. Pesquisamos, sob a ótica do critério AIC, qual seria o melhor ajuste, considerando todas as combinações das ordens $p, q, m, s \leq 5$, e também várias outras extensões do modelo como a existência de memória longa na média e na volatilidade, do parâmetro de *leverage*, do parâmetro de prêmio pelo risco, e ainda duas possibilidades para a distribuição dos erros, a Normal e t-student. Lembramos que melhor modelo ajustado deve ter todos os parâmetros significativos (consideramos aqui o nível de significância de 5%).

Ao nível de significância de 5% nenhuma dependência foi achada para a média, isto é, $p = q = d = 0$, sendo d o parâmetro de memória longa. Os log-retornos são então um ruído branco.

O teste de Ljung-Box aplicado aos quadrados dos retornos e o teste de ML aplicados aos retornos, ambos para 5 *lags*, rejeitaram fortemente suas respectivas hipóteses nulas de “as autocorrelações de lag 1 a 5 são iguais a zero” e de “não existe efeito ARCH(5)” com valores da estatística teste e do p-valor sendo, respectivamente, 755.99 e 0.0000, e 424.76 e 0.0000.

A seguir procuramos pelo melhor ajuste GARCH(m, s) com seus adicionais, e também pelo ajuste básico (*benchmark*) GARCH(1,1) Gaussiano, e pelo melhor ajuste FIGARCH(m, d, s). A Tabela 5.2 mostra os resultados da estimação por máxima verossimilhança. Observamos que o melhor ajuste (m, s) continua sendo (1, 1) sem o parâmetro *leverage*, mas que a suposição da distribuição condicional t melhora consideravelmente o ajuste, diminuindo muito o AIC. Apesar de ser fato comumente observado para séries de retornos de ações, aqui vemos que o impacto de uma inovação negativa *não* é maior do que aquele observado quando de um retorno positivo. Incluindo memória longa na volatilidade não melhora o ajuste, sendo o AIC deste modelo maior do que o modelo anterior, apesar do parâmetro de memória longa ser estatisticamente significativo. Note entretanto que este ajuste FIGARCH é baseado na distribuição normal e seu AIC é menor que o GARCH(1,1) Gaussiano.

Para os três ajustes o teste de Ljung-Box com 12 defasagens aplicado aos quadrados dos resíduos padronizados aceitou a hipótese nula de *não autocorrelação* com p-valores respectivamente iguais a 0.7726, 0.7029, 0.9497. Assim, os três modelos capturaram adequadamente a persistência na variância dos retornos, fato confirmado pela inspeção visual da f.a.c.a. da série dos resíduos padronizados ao

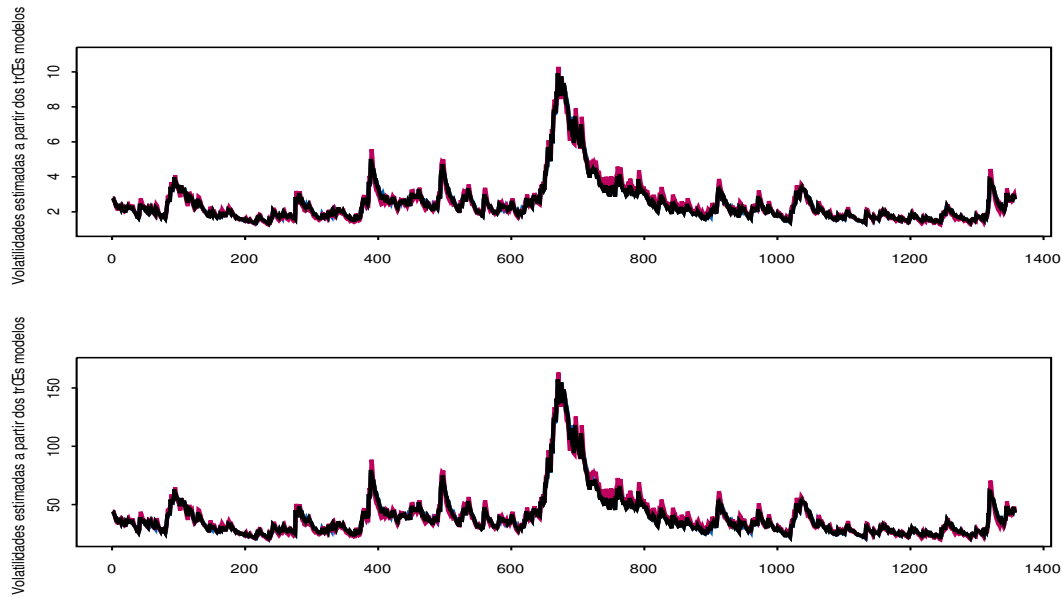


Figura 5.3: As volatilidades estimadas a partir dos três modelos. A linha preta corresponde ao GARCH(1,1). A rosa ao modelo com memória longa, a qual parece produzir os maiores valores para a volatilidade. Na parte superior a diária, e na parte inferior a anualizada.

quadrado. Para 20 lags os p-valores são, respectivamente, 0.7842, 0.6899, 0.9114. Gostaríamos de ressaltar que estamos fazendo todos os ajustes e cálculos utilizando o SPlus, e que os valores são ligeiramente diferentes daqueles fornecidos pelo R e pelo E-views.

A Figura 5.3 mostra as volatilidades estimadas a partir dos três modelos. Na parte superior a diária, e na parte inferior a anualizada. É muito interessante observar que visualmente aquelas provenientes dos dois modelos *sem* memória longa são praticamente indistinguíveis. A linha preta corresponde ao GARCH(1,1) Gaussiano. A rosa, fornecida pelo modelo com memória longa, parece produzir os maiores valores para a volatilidade.

Para o modelo GARCH(1,1) a “persistência” é igual a $(\alpha_1 + \beta_1)$ que fornece os valores 0.9797 e 0.9835 para os dois primeiros modelos. Sendo $(\alpha_1 + \beta_1)$ estatisticamente menor que um (teste t unilateral para a direita rejeita a 5% a hipótese nula de que a soma é um), podemos admitir que a volatilidade reverte para a média, apesar da meia vida ser relativamente alta. De fato, para os dois primeiros modelos a meia vida da volatilidade é igual a 35 e 43 dias, respectivamente.

A variância não condicional dos retornos, de acordo com o modelo GARCH(1,1) é dada por $\gamma_0 = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}$. Para os modelos M1 e M2 é igual a 6.73 e 6.78, respectivamente. Em geral esperamos um valor maior para o modelo Gaussiano devido a falta de robustez dos estimadores de máxima verossimilhança o que faz com que a estimativa da variância fique inflada. Esses valores implicam, para esta amostra, numa volatilidade (desvio padrão) média anualizada de 41.19% e de 41.03%,

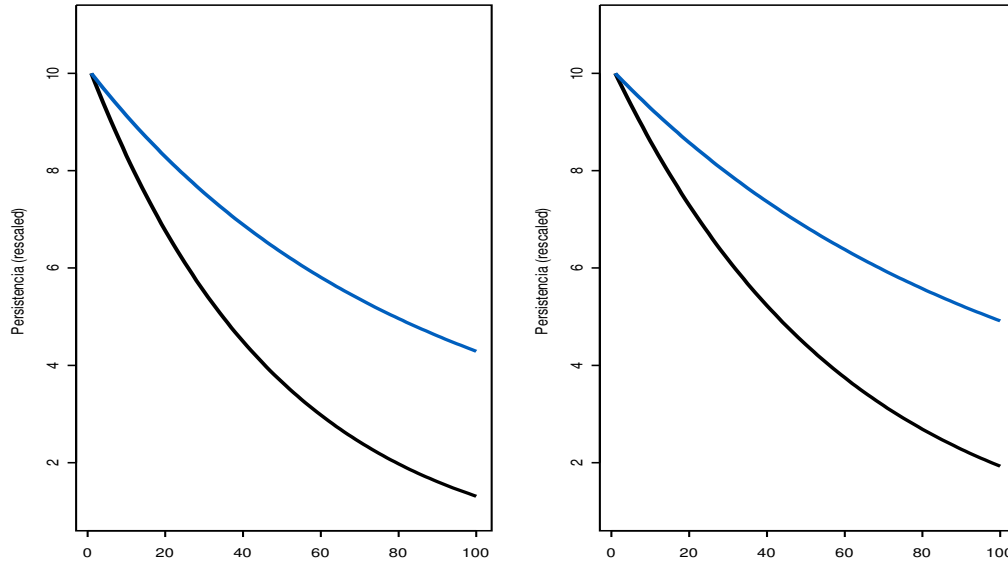


Figura 5.4: As persistências a frente $\theta_{t+k|t}$ e acumulada $\phi_{t+k|t}$ estimadas a partir dos dois modelos sem memória longa M1 (esquerda) e M2 (direita). Em cada gráfico a linha preta corresponde a persistência para a frente, e a linha azul representa a persistência acumulada. As curvas estão multiplicadas por um fator.

lembrando que aquela calculada a partir do desvio padrão amostral foi de 44.37%. Lembramos que, por sua vez, o desvio padrão amostral não é um estimador robusto, possuindo ponto de ruptura zero.

Na Figura 5.4 mostramos as *forward persistence* $\theta_{t+k|t}$ para $k = 1, \dots, 100$ calculadas para os modelos M1 (a esquerda) e M2 (a direita). Notemos que as quantidades $\theta_{t+k|t}$ estão multiplicadas por $\frac{10}{\theta_{t+1|t}}$. As curvas decaem geometricamente, sendo seu ultimo termo (sem o fator de escala) muito pequeno (0.0130 e 0.0176), mostrando que o impacto da informação no tempo t se esvai quase por completo após 100 períodos. Após dois anos é praticamente zero, indicando a reversão da volatilidade para a média. Esta figura também mostra as persistências GARCH(1,1) acumuladas $\phi_{t+k|t}$, em azul e multiplicadas por um fator de escala. As curvas para os dois modelos também decrescem para zero porém mais lentamente.

Uma outra maneira de se observar a reversão para a média da volatilidade GARCH é através do gráfico de suas previsões a longo prazo. A Figura 5.5 mostra previsões feitas a partir de duas datas e de 1 até 200 passos a frente, para a volatilidade (desvio padrão condicional) anualizada do retorno diário. Uma das datas está num período de baixa volatilidade (27 de janeiro de 2001) e a outra em um período de alta volatilidade (31 de outubro de 2008). Novamente mostramos a esquerda e direita os resultados para os dois primeiros modelos estimados. Lembrando que os níveis médios são, respectivamente, 41.19% e de 41.03%, observamos na figura que, de um modelo para o outro, embora

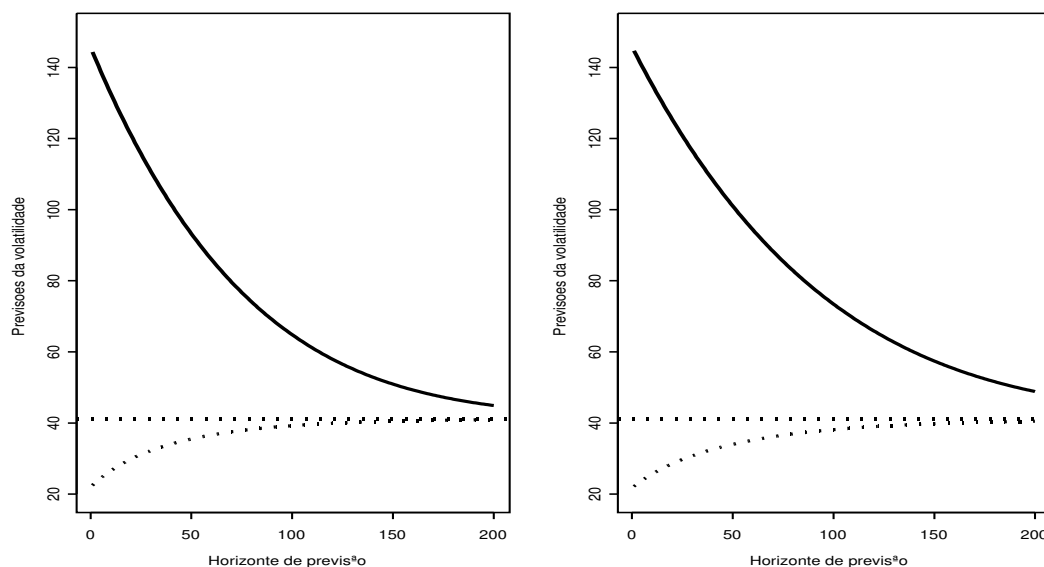


Figura 5.5: Previsões da volatilidade anualizada para 200 passos a frente. Uma das datas de início das previsões corresponde a um período de baixa volatilidade (27 de janeiro de 2001) e a outra em um período de alta volatilidade (31 de outubro de 2008). A esquerda e direita os resultados para os modelos M1 e M2.

as previsões sejam bem similares, produzem figuras bem diferentes devido ao fato da persistência sob M2 ser maior, assim o decaimento para o modelo M1 é mais acentuado.

Calculamos agora a VoV , dada como o desvio padrão do retorno da variância, $rv_t = \frac{\sigma_t^2 - \sigma_{t-1}^2}{\sigma_{t-1}^2}$. Para os três modelos a VoV calculada para a variância condicional um passo a frente é, respectivamente igual a 0.1748, 0.1623, e 0.1590. Anualizando este desvio padrão temos 2.775, 2.576, e 2.524. Uma outra maneira de inspecionar a VoV sugerida em Engle, R. F. e Patton, A. J. (2001) é comparar os gráficos da volatilidade um passo a frente e da previsão da volatilidade k -passos a frente. A Figura 5.6 mostra essas previsões acumuladas para 1 passo a frente e 63 dias. Conforme esperado, a medida que o horizonte se afasta mais suave são os movimentos da curva.

O desvio padrão anualizado das previsões GARCH um passo a frente anualizadas, isto é, a VoV anualizada dos desvios padrões condicionais previstos um passo a frente anualizados para os três modelos estimados são respectivamente iguais a 307, 310, e 315, sob os três modelos. Abaixo damos esta VoV anualizada para os dois modelos sem memória longa e para dois cenários de previsão, 1 e 63 dias.

	Um dia	3 meses
GARCH(1,1) Gaussiano	307	111
GARCH(1,1) t_5	310	136

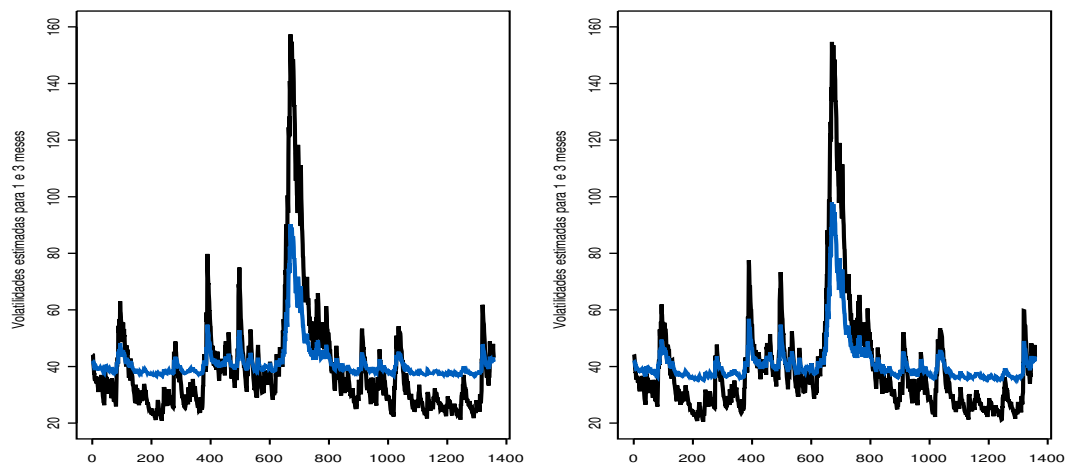


Figura 5.6: *Previsões anualizadas da volatilidade 1 passo a frente e 63 dias e para os dois modelos (esquerda e direita).*

5.5 Volatilidade Realizada

A crescente disponibilidade de dados de alta frequência para vários tipos de instrumentos financeiros e em todos os mercados, têm motivado o uso da volatilidade realizada e de outras medidas de variabilidade realizada. A volatilidade realizada (RV) é um estimador não viciado e eficiente da volatilidade, que não é estimada a partir de um modelo, mas simplesmente calculada a partir de dados de alta frequência. Para qualquer dia fixo t , a RV_t é definida como a soma dos quadrados dos retornos intra diários (veja detalhes teóricos em Andersen (1998), Andersen, T. G., Bollerslev, T., Diebold, F. X. e Labys, P. (2001), Andersen, T. G., Bollerslev, T., Diebold, F. X. e Ebens, H. (2001), e Andersen, T. G., Bollerslev, T., Christoffersen, P. F. e Diebold, F. X. (2006). Este termo foi usado pela primeira vez por Fung, W. K. H. e Hsieh, D. A. (1991) e Andersen, T. G. e Bollerslev, T. (1998b) como uma proxy da volatilidade.

Dados de alta frequência têm-se tornado muito mais disponíveis recentemente, dando origem ao que se passou a chamar de “Big Data era”. O processamento desta quantidade massiva de dados é mais complexo e demorado e passou a requerer melhores pacotes computacionais e maior conhecimento da parte dos programadores.

Dados de alta frequência possuem características únicas que não estão presentes em dados de baixa frequência (diários, semanais, mensais) e necessitam de tratamento especial. Transações ocorrem em intervalos de tempo irregulares e observamos atividade de mercado variável ao longo do dia. Assim, os dados podem apresentar sazonalidades, falta de sincronicidade, serem afetados pela situação de *bid-ask* e tudo isto pode finalmente distorcer as inferências feitas. Além do mais, o número de observações é grande, aumentando as chances de vários tipos de erros, por exemplo, várias formas de erros de

registros de transações. Mas tudo isto apenas torna as análises estatísticas mais interessantes.

Aplicações da volatilidade realizada são várias. Em Andersen, T. G. (2000) e Andersen, T. G., Bollerslev, T., Diebold, F. X. e Labys, P. (2003) temos evidências de que o uso de retornos de alta frequência melhora muito a previsão da volatilidade para horizontes de um dia ou mesmo de até um mês. A quantificação do risco de mercado intra-diário ajuda os participantes de mercado, como por exemplo os *traders*. So, M. K. P. e Xu, R. (2013) sugerem que a inclusão de índices intra-diários na volatilidade GARCH captura outras informações em mercados muito voláteis.

Segundo alguns autores (por exemplo, Andersen, T. G., Bollerslev, T. e Diebold, F. X. (2007)), os processos dos log-preços seriam mais bem descritos como uma combinação de um processo suave, contínuo, de reversão para a média bem lenta e de um outro menos persistente, baseado em *jumps*. Veja por exemplo Barndoff Nielsen, O. E. e Shephard, N. (2006a,b), onde são estudadas as medidas de variação *bipower*.

Seja $r(t) = p(t) - p(0)$. Definimos a variação quadrática do processo de retornos $r(t)$, $[r, r]_t$, como

$$[r, r]_t = \int_0^t \sigma^2(s) ds \quad (5.13)$$

onde o lado direito é a volatilidade integrada do componente contínuo do processo de difusão de tempo contínuo, usualmente expresso pela equação diferencial estocástica como

$$dp(t) = \mu(t)dt + \sigma(t)dW(t), \quad 0 \leq t \leq T \quad (5.14)$$

onde $\mu(t)$ é um processo contínuo e localmente limitado, $\sigma(t)$ é um processo de volatilidade estocástica estritamente positivo, $W(t)$ um movimento Browniano.

Salto podem ser incorporados no processo do log-preço:

$$dp(t) = \mu(t)dt + \sigma(t)dW(t) + \mathcal{K}(t)dq(t) \quad (5.15)$$

onde $q(t)$ é um processo de contagem com intensidade que pode variar no tempo e onde $\mathcal{K}(t) = p(t) - p(t^-)$ representa os saltos do processo.

Considere o processo de retornos amostrados a cada Δ -períodos, $r_{r,\Delta} \equiv p(t) - p(t - \Delta)$. Para simplificar a notação, iremos normalizar o intervalo de tempo diário como sendo 1. Assim, o retorno diário tem $\Delta = 1$ e o denotaremos simplesmente por r_{t+1} em vez de $r_{t+1,1}$. A variação, ou volatilidade, diária realizada (RV), é calculada como a soma dos quadrados dos $\frac{1}{\Delta}$ retornos de alta frequência.

$$RV_{t+1}(\Delta) \equiv \sum_{j=1}^{\frac{1}{\Delta}} r_{t+j\cdot\Delta,\Delta}^2 \quad (5.16)$$

onde, sem perda de generalidade, assumimos ser $\frac{1}{\Delta}$ inteiro. Veja Andersen, T. G. e Bollerslev, T. (1998a), Andersen, T. G., Bollerslev, T., Diebold, F. X. e Labys, P. (2001), Andersen, T. G., Bollerslev, T., Diebold, F. X. e Labys, P. (2003), entre outros. Dados intra diários apresentam sazonalidades e clusters de volatilidade. Para eliminar o ruído devido aos efeitos de microestrutura um filtro ARFIMA(p, d, q) pode ser usado antes de se calcular a volatilidade realizada.

Conforme provado (Andersen, T. G. e Bollerslev, T. (1998a), Andersen, T. G., Bollerslev, T., Diebold, F. X. e Labys, P. (2001)), a RV converge uniformemente em probabilidade para o incremento do processo de variação quadrática, quando a frequência da amostragem cresce ($\Delta \rightarrow 0$).

$$RV_{t+1}(\Delta) \xrightarrow{\Delta \rightarrow 0} \int_t^{t+1} \sigma^2(s) ds + \sum_{t < s \leq t+1} \mathcal{K}^2(s) \quad (5.17)$$

Assim, a RV deve refletir as dinâmicas de outros processos (contínuo e de saltos). Logo, melhores previsões podem ser feitas se incorporarmos saltos na modelagem da RV. Veja detalhes desta modelagem (*realized bipower variation*), $BV_{t+1}(\Delta)$, em Barndoff Nielsen, O. E. e Shephard, N. (2006a,b).

$$BV_{t+1}(\Delta) \equiv \mu_1^{-2} \sum_{j=2}^{\frac{1}{\Delta}} |r_{t+j\Delta, \Delta}| \cdot |r_{t+(j-1)\Delta, \Delta}| \quad (5.18)$$

onde $\mu_1 = \sqrt{\frac{2}{\pi}} = E(|Z|)$.

Barndoff Nielsen, O. E. e Shephard, N. (2006a), mostram que a contribuição para o processo de variação quadrática devido ao processo de saltos (descontinuidades), pode ser consistentemente estimada, quando $\Delta \rightarrow 0$, por

$$RV_{t+1}(\Delta) - BV_{t+1}(\Delta) \longrightarrow \sum_{t < s \leq t+1} \mathcal{K}^2(s) \quad (5.19)$$

onde $BV_{t+1}(\Delta)$ é calculado como em (5.18). Andersen, T. G., Bollerslev, T. e Diebold, F. X. (2007) sugerem minimizar os efeitos de microestrutura usando retornos de 5-minutos para mercados bastante ativos. Uma outra possibilidade para aliviar os efeitos de microestrutura é usar variação realizada *staggered*.

Na prática, para evitar que o lado direito de (5.19) seja negativo, os autores sugerem calcular o *salto realizado* (JR) como:

$$J_{t+1}(\Delta) = \max(RV_{t+1}(\Delta) - BV_{t+1}(\Delta), 0) \quad (5.20)$$

Veja exercício 5.1 no final deste capítulo onde sugerimos uma análise comparativa entre essas medidas de volatilidade realizada e a volatilidade GARCH.

Modelos HAR-RV

Seja

$$RV_{t,t+h} = \frac{1}{h} (RV_{t+1} + RV_{t+2} + \dots + RV_{t+h}), \quad h = 1, 2, \dots \quad (5.21)$$

a variância realizada de multiperíodos normalizada. Note que $RV_{t,t+1} = RV_{t+1}$, e que $E[RV_{t,t+h}] = E[RV_{t+1}] \quad \forall h$. Em geral tomamos $h = 5$ e $h = 22$, os quais correspondem às volatilidades semanais e mensais.

O modelo de Corsi, F. e Curci, G. (2003) é dado por

$$RV_{t+1} = \beta_0 + \beta_D RV_t + \beta_W RV_{t-s,t} + \beta_M RV_{t-22,t} + \varepsilon_{t+1} \quad (5.22)$$

$t = 1, 2, \dots, T$.

Notemos que volatilidades relacionadas com outros h , poderiam ser incluídas na regressão acima. O modelo pode ser ajustado por MQ para obter as estimativas $\widehat{\beta}_0$, $\widehat{\beta}_D$, $\widehat{\beta}_W$ e $\widehat{\beta}_M$ os quais devem ser todos significativos. Sempre observe o R^2 .

A primeira extensão óbvia é para horizontes maiores de previsão.

$$RV_{t,t+h} = \beta_o + \beta_0 RV_t + \beta_W RV_{t-s,t} + \beta_M RV_{t-22,t} + \varepsilon_{t,t+h}.$$

Andersen, T. G., Bollerslev, T. e Diebold, F. X. (2007) sugerem incluir a série de saltos J_t como variável explanatória

$$RV_{t,t+h} = \beta_o + \beta_0 RV_t + \beta_W RV_{t-s,t} + \beta_M RV_{t-22,t} + \beta_J J_t + \varepsilon_{t,t+h} \quad (5.23)$$

Sugerimos estimar o modelo (5.23) para $h = 1$ para as séries (v. dependente) RV_t , $RV_t^{\frac{1}{2}}$ e $\log(RV_t)$ de algum ativo como exercício.

Modelo HAR-RV

A RV no tempo t pode ser modelada por um modelo de regressão com constante e tendo como variáveis explicativas a RV no tempo $t - 1$, a RV semanal, e a mensal. Ver definição e modelo acima. Claro que podemos listar outras variáveis explicativas, como por exemplo a RV com $h = 2$, $h = 3$ ou $h = 10$ (2 semanas).

Notemos que as séries de RV devem ser estacionárias. Os testes de estacionaridade e de RU vistos nos capítulos 1 e 4, devem ser aplicados. Uma estimativa do grau de integração fracionária, d , pode ser obtida com os estimadores baseados no log-periodograma (Mendes, B. V. M. e Lopes, S. R. C. (2006)).

A novidade trazida por Andersen, T. G., Bollerslev, T. e Diebold, F. X. (2007) é a inclusão do processo de saltos no modelo de regressão de Corsi, F. e Curci, G. (2003), dando origem ao modelo HAR-RV-J

$$RV_{t,t+h} = \beta_o + \beta_0 RV_t + \beta_W RV_{t-s,t} + \beta_M RV_{t-22,t} + \beta_J J_t + \varepsilon_{t,t+h}.$$

Devido a própria definição das quantidades $RV_{t,t+h}$ fica claro que os erros serão correlacionados. Isto não afeta a consistência dos estimadores, mas afeta seus erros padrões.

Novamente sugerimos ao leitor ajustar o modelo e obter as estimativas e erros padrões de β_0 , β_W , β_M , Se significativos, indicam persistência. Observe que a estimativa de β_J pode ser negativa, e devemos ter em mente a possibilidade das previsões serem negativas. Sugerimos também comparar os ajustes das regressões com e sem saltos e comparar os R^2 's das duas regressões. Observe se a inclusão dos saltos faz com que os outros coeficientes (as RV's passadas) percam importância, e verifique se de fato o salto é a variável mais importante na previsão.

Algumas outras observações: (1) Costuma-se estimar o modelo HAR-RV-J para os desvios padrões. Por exemplo, o lado esquerdo seria $(RV_{t,t+h})^{\frac{1}{2}}$. (2) O $\log(RV_{t,t+1})$ poderia também ter sido usado. (3) Poderia comparar também com o modelo ARFIMA(p, d, q).

Os saltos foram definidos como a diferença entre a variação *bipower* e a RV, sendo assim sujeitos a erros. Existe uma corrente que defende identificar os saltos pequenos com esses erros de mensuração e associar apenas os grandes valores de $RV_t(\Delta) - BV_t(\Delta)$ como um salto. Neste caso, saltos significativos podem ser identificados através de um teste formal, cuja estatística teste é dada em Andersen, T. G., Bollerslev, T. e Diebold, F. X. (2007), seguindo resultados em Barndorff Nielsen, O. E. e Shephard, N. (2006a). Esta modificação pode ser incluída na regressão e em geral melhora os resultados. O teste de Ljung-Box pode ser aplicado nas séries de saltos. Os saltos (pequenos e grandes) podem aparecer (ou não) em conglomerados.

É sempre interessante verificar se a dinâmica dos saltos são menos persistentes e previsíveis que a dinâmica da trajetória contínua. Esta medida de saltos, baseada na variação realizada, calculada a partir de dados de alta frequência, permite identificar mais saltos do que os modelos existentes (anteriores) baseados em dados diários. Outra coisa é que a maioria dos saltos significativos (estatisticamente significantes), estão associados com anúncios de notícias macroeconômicas. Alguns autores empiricamente observaram que no modelo de regressão contínua incluindo como variáveis explicativas, a trajetória contínua, e os saltos, apenas a trajetória contínua tinha poder preditivo, e que este modelo era melhor do que os existentes na literatura baseados também na RV. Mesmo assim, é interessante separar os dois movimentos para o gerenciamento do risco associado à eles.

Outras definições da RV

A frequência de amostragem, obviamente, deve ser maior ou igual a de transação-em-transação. A questão da escolha da frequência ótima foi abordada em diversos trabalhos, veja por exemplo, Zhou, B. (1996), Andersen et al. (2001a), Andersen et al. (2001b), Andersen, T. G., Bollerslev, T., Diebold, F. X. e Labys, P. (2003), Barndorff-Nielsen and Shephard (2002), Barndorff-Nielsen, O. E. e Shephard, N. (2002), e Meddahi, N. (2002), e a maioria dos trabalhos aponta como sendo de 5 a 30 minutos este intervalo ótimo. Entretanto, dado um intervalo, perde-se observações potencialmente importantes. Estudos mais recentes tais como Zhang et al. (2005), Aït-Sahalia, Y. (2005), Hansen, P. R. e Lunde, A. (2006), Bandi, F. M. e Russell, J. R. (2008), Aït-Sahalia, Y., Mykland, P. A. e Zhang, L. (2011), e Barndorff-Nielsen et al. (2008, 2011) usam (quase) toda a informação disponível para estimar a volatilidade, dando origem aos estimadores denominados de *variância realizada robusta*, que apesar de chamada de *robusta*, não seria resistente aos ruídos, sendo um estimador ruim para a variância integrada. Seja \underline{h} um intervalo de amostragem muito pequeno. Este estimador é definido como

$$RV_t^{all}(\underline{h}) \equiv RV_t(\underline{h}). \quad (5.24)$$

Seja h um intervalo de amostragem múltiplo de \underline{h} . Isto é, $h = \underline{h}n_h$, onde n_h é um inteiro positivo, e $n_h - 1$ representa o número de observações ignoradas em cada intervalo amostral. O estimador *esparso*, $RV_t(h)$, para h grande, é mais robusto em relação ao ruído

$$RV_t^{sparse}(h) \equiv RV_t(h). \quad (5.25)$$

Este estimador não considera $n_h - 1$ informações para cada h intervalo amostral. Aït-Sahalia, Y. (2005) propôs o RV médio, que utiliza todos os dados, como a média de n_h estimadores esparsos. O

estimador esparsos com janela de amostragem deslocada de $k\mathbf{h}$ é definido como

$$RV_t^{sparse}(h, k) = \sum_{j=1}^{N_k} r_{t-1+jh+k\mathbf{h}}^2, \quad k = 0, \dots, n_h - 1, \quad (5.26)$$

onde

$$N_k = \frac{1}{h} \text{ if } k = 0, \quad N_k = \frac{1}{h} - 1 \text{ if } k = 1, \dots, n_h - 1. \quad (5.27)$$

O estimador médio é então definido como a média dos estimadores esparsos (5.26),

$$RV_t^{average}(h) = \frac{1}{n_h} \sum_{k=0}^{n_h-1} RV_t^{sparse}(h, k). \quad (5.28)$$

Existem inúmeras outras definições de medidas de variação realizada, veja Zhang, L. (2006), Zhou, B. (1996), Barndorff-Nielsen, O. E. e Shephard, N. (2004), entre vários outros. O estimador que pondera pelo volume de transação, o *volume-weighted average price* (VWAP) é muito usado como referência, embora seja bastante popular o *closest*, que utiliza a mais recente observação, ou mesmo aquele que usa algum tipo de interpolação dos preços.

Ting, C. (2006) e Mendes, B. V. M. e Accioly, V. B. (2012) fornecem evidências de que o VWAP produz valores menores para a RV, podendo assim serem chamados de robustos. Sejam P_j e Q_j , respectivamente, os j -ésimos preço e volume negociados por k vezes de um ativo durante um intervalo de tempo Δ . O log-preço $p_{t-1+\Delta}$ de $t-1$ a $t-1+\Delta$ é definido como

$$p_{t-1+\Delta}^{vwap} = \ln \left(\frac{P_1 \cdot Q_1 + P_2 \cdot Q_2 + \dots + P_k \cdot Q_k}{Q_1 + Q_2 + \dots + Q_k} \right). \quad (5.29)$$

Os log-retornos intra diários são obtidos a partir desses preços ponderados.

O Range Realizado (RR)

Em Martens, M. e van Dijk, D. (2007) e Christensen, K. e Podolskij, M. (2007) temos o desenvolvimento teórico de um novo estimador chamado de *realized range*, RR, que evita o problema da dependência do tamanho do intervalo amostral. Por exemplo, Parkinson, M. (1980) e Martens, M. e van Dijk, D. (2007) definem o RR como

$$RR_t^{Parkinson}(h) = \frac{1}{4 \ln 2} \sum_{j=1}^{1/h} (u_j - d_j)^2, \quad (5.30)$$

$$u_j = \ln H_j - \ln O_j \quad d_j = \ln L_j - \ln O_j, \quad (5.31)$$

onde H_j , L_j e O_j são os preços maior, menor, e de abertura do j -ésimo intervalo no t -ésimo dia útil; e u_j e d_j são preços alto e baixo normalizados. Para mais detalhes veja também Martens, M. e van Dijk, D. (2007), Garman, M. B. e Klass, M. J. (1980), Rogers, L. C. G. e Satchell, S. E. (1991), Yang, D. e Zhang, Q. (2000) e Rogers, L. C. G. e Satchell, S. E. (1991).

5.6 Exercícios

Exercício 1 O objetivo deste exercício é melhorar as previsões da volatilidade. Obtenha dados de preços intra diários baseados nas transações *tick-by-tick* e construa os retornos 5-minutos a partir de uma interpolação linear logarítmica. Fazer o gráfico dos preços de 5-minutos no tempo *para algum dia* onde se observe salto grande e para *um outro dia* sem salto grande. Identifique dia e hora. Aconteceu algo neste dia/horário?

Considere também os retornos *overnight* de acordo com as opções vistas. Teremos um total de $\frac{1}{\Delta} = 98$ intervalos de 5-minutos por dia. Daqui para frente nos referiremos às equações (5.16) e (5.20) respectivamente como RV_t e J_t . Iremos considerar as 6 séries RV_t , $RV_t^{\frac{1}{2}}$, $\log(RV_t)$, J_t , $J_t^{\frac{1}{2}}$, e $\log(J_t + 1)$.

(a) Inicie examinando a distribuição não condicional das 6 séries, por exemplo examinando o histograma delas. Notemos que em Mendes, B. V. M. e Accioly, V. B. (2012) foi observado que a série $\log(RV_t^{\frac{1}{2}})$ apresenta distribuição mais próxima da Normal, sendo por isto mais utilizada. Além disto, usar o log evita a questão da não negatividade. Calcule e compare as médias das séries J_t e RV_t .

(b) Faça um gráfico das 6 séries no tempo.

(c) Examine a f.a.c. amostral das 6 séries. Provavelmente irão mostrar autocorrelação significativa para vários *lags*.

(d) Realize os testes de estacionariedade e de raiz unitária para as séries. Você rejeita a H_0 de raiz unitária? Quais foram os p-valores? Você observa que as maiores volatilidades realizadas estão associadas aos saltos na série dos log-preços? Anote as datas dos maiores saltos.

(e) Calcule a estatística de Ljung-Box (LB) com $m = 12$ para as 6 séries. Compare os valores da estatística $LB(m=12)$ para as séries J_t , $J_t^{\frac{1}{2}}$ e $\log(J_t + 1)$, com os valores obtidos para as séries RV . Que conclusão pode tirar a respeito da localização da maior parte da dinâmica temporal da série de variação quadrática? Seria devida à parte contínua RV_t ? Ou aos saltos?

(f) Consulte e implemente as idéias em Andersen, T. G., Bollerslev, T. e Diebold, F. X. (2007), onde os autores sugerem um modelo para previsão da volatilidade baseado num modelo previamente proposto em Andersen, T. G., Bollerslev, T., Diebold, F. X. e Labys, P. (2003). A idéia é bem simples. Consiste em usar as séries J_t , $J_t^{\frac{1}{2}}$ e $\log(J_t)$ como variáveis explicativas em um modelo de regressão.

(g) Ajuste um modelo ARFIMA na série $RV_t^{\frac{1}{2}}$. Observe novamente a f.a.c. das séries RV_t , $RV_t^{\frac{1}{2}}$ e $\log(RV_t)$. Elas sugerem memória longa? Mendes, B. V. M. e Accioly, V. B. (2012) comentam a adequação dos modelos ARFIMA para capturar os efeitos de microestrutura.

(h) Finalmente, assim como em Andersen, T. G., Bollerslev, T. e Diebold, F. X. (2007), estime o modelo HAR-RV de Corsi, F. e Curci, G. (2003). Os modelos HAR-RV seriam uma extensão dos HARCH (modelos ARCH Heterogêneos de Muller et. al. (1997)), os quais, apesar de não modelarem a memória longa explicitamente, possuem uma f.a.c. com decaimento lento muito semelhante ao padrão de memória longa.

(i) Finalmente utilize os ajustes obtidos para fazer previsões da volatilidade e calcular alguma medida de risco, por exemplo o VaR. Faça o exercício *out-of-sample* por um período, digamos de 6 meses, e verifique qual previsão da volatilidade produz melhores estimativas do VaR.

Exercício 2 Faça os vários exercícios sugeridos na Seção 5.5.

Capítulo 6

Modelos Condicionais Multivariados

Este capítulo apresenta a versão *multivariada* dos modelos dinâmicos (*condicionais*) univariados dados para a média (Capítulo 4, ARFIMA), e para a estrutura de dependência (Capítulo 3, cópulas).

6.1 Modelos Vetoriais Autorregressivos (VAR(p))

O modelo VAR (*vector autoregressive*) é a extensão natural para séries temporais multivariadas estacionárias do modelo univariado AR(p) visto no Capítulo 4. Seja $\mathbf{x}_t = (x_{1,t}, x_{2,t}, \dots, x_{d,t})'$ um vetor aleatório ($d \times 1$). Modelo VAR(p) pode ser escrito como

$$\mathbf{x}_t = \mathbf{c} + \mathbf{\Pi}_1 \mathbf{x}_{t-1} + \mathbf{\Pi}_2 \mathbf{x}_{t-2} + \dots + \mathbf{\Pi}_p \mathbf{x}_{t-p} + \boldsymbol{\varepsilon}_t, \quad t = 1, \dots, T \quad (6.1)$$

onde \mathbf{c} é um vetor ($d \times 1$) de constantes, $\mathbf{\Pi}_i$ são matrizes ($d \times d$) de coeficientes e $\boldsymbol{\varepsilon}_t$ é um vetor ($d \times 1$) de erros não observáveis, gerados a partir de um processo RB com média zero e matriz de covariâncias $\boldsymbol{\Sigma}$ constante no tempo. Notemos que temos d equações com os mesmos regressores no lado direito.

Considerando um VAR(1) estacionário temos que a esperança não condicional do processo é

$$E(\mathbf{x}_t) = \mathbf{c}(\mathbf{I}_d - \mathbf{\Pi}_1)^{-1}$$

desde que $(\mathbf{I}_d - \mathbf{\Pi}_1)$ seja não singular. A variância não condicional é $var(\mathbf{x}_t) = \sum_{i=0}^{\infty} \mathbf{\Pi}_1^i \boldsymbol{\Sigma} (\mathbf{\Pi}_1^i)'$, onde $\mathbf{\Pi}_1^0 = \mathbf{I}_d$, e a covariância de lag k é dada por $cov(\mathbf{x}_t, \mathbf{x}_{t-k}) \equiv cov_k = \mathbf{\Pi}_1 cov_{k-1}$, $k > 0$, uma generalização do resultado visto para o processo AR(1).

Escrito de uma maneira compacta com o operador retroativo, o VAR(p) é

$$\mathbf{\Pi}(B)\mathbf{x}_t = \mathbf{c} + \boldsymbol{\varepsilon}_t \quad (6.2)$$

onde $\mathbf{\Pi}(B) = \mathbf{I}_d - \mathbf{\Pi}_1 B - \dots - \mathbf{\Pi}_p B^p$.

O VAR(p) é estável se as raízes de $\det(I_d - \Pi_1 Z - \dots - \Pi_p Z^p) = 0$ estão fora do círculo unitário complexo (isto é, possuem valor absoluto maior que 1). Neste caso o processo VAR(p) é estacionário de 2ª ordem e ergódico. A média não condicional é $\mu = (I_d - \Pi_1 - \dots - \Pi_p)^{-1}c$. A $\text{cov}(\mathbf{x}_t, \varepsilon_t) = \Sigma$ (a matriz de covariâncias de ε_t), $\text{cov}(\mathbf{x}_{t-k}, \varepsilon_t) = 0$ para $k > 0$, e $\text{cov}_k = \Pi_1 \text{cov}_{k-1} + \dots + \Pi_p \text{cov}_{k-p}$ para $k > 0$. Notemos que a f.a.c. e a f.a.c.p. vistas no caso univariado, são também facilmente estendidas, veja Tsay, R. S. (2005). O mesmo ocorre com a generalização para os modelos VMA(q) e os VARMA(p, q).

Podemos incluir no modelo VAR(p) termos determinísticos como uma tendência linear no tempo, ou variáveis *dummies* sazonais, além de variáveis exógenas \mathbf{y}_t . A forma geral é:

$$\mathbf{x}_t = \Pi_1 \mathbf{x}_{t-1} + \Pi_2 \mathbf{x}_{t-2} + \dots + \Pi_p \mathbf{x}_{t-p} + \Phi \mathbf{D}_t + \mathbf{G} \mathbf{y}_t + \varepsilon_t, \quad (6.3)$$

onde \mathbf{D}_t é uma matriz ($l \times 1$) de componentes determinísticos, Φ e \mathbf{G} são matrizes de parâmetros, e \mathbf{y}_t é uma matriz ($m \times 1$) de variáveis exógenas.

Para a estimação do VAR(p), supondo que não haja restrições nos parâmetros, considere as equações (de regressão)

$$\mathbf{x}_i = \mathbf{Z} \pi_i + \mathbf{e}_i, \quad i = 1, \dots, d \quad (6.4)$$

onde \mathbf{x}_i é um vetor ($T \times 1$) de observações da i -ésima equação, \mathbf{Z} é uma matriz ($T \times k$), onde a t -ésima linha é $\mathbf{Z}'_t = (1, \mathbf{x}'_{t-1}, \dots, \mathbf{x}'_{t-p})$ e portanto $k = dp + 1$, π_i é um vetor ($k \times 1$) de parâmetros e \mathbf{e}_i é um vetor ($T \times 1$) de erros com matriz de covariâncias $\sigma_i^2 \mathbf{I}_T$. Como todas as d equações possuem as mesmas variáveis explanatórias, elas podem ser estimadas separadamente por MQO sem perder eficiência. Cada estimador $\hat{\pi}_i$ tem distribuição assintótica Normal com erro padrão dado pela raiz quadrada do i -ésimo elemento da diagonal de $\hat{\Sigma} \otimes (\mathbf{Z}'\mathbf{Z}^{-1})$, onde $\hat{\Sigma}$ é a matriz de covariâncias dos resíduos de mínimos quadrados $\hat{\varepsilon}_i$ de (6.1).¹

¹O modelo VAR(p) pode ser reescrito em forma compacta como

$$\Pi(B) \mathbf{x}_t = \Phi \mathbf{D}_t + \varepsilon_t$$

onde $\Pi(B) = I_d - \Pi_1 B - \dots - \Pi_p B^p$ é uma matriz polinomial, ou, equivalentemente

$$\mathbf{x}_t = \mathbf{B}' \mathbf{Z}_t + \varepsilon_t$$

onde $\mathbf{B}' = (\Pi_1, \Pi_2, \dots, \Pi_p, \mathbf{c})$ e $\mathbf{Z}_t = (\mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \dots, \mathbf{x}_{t-p}, 1)$, assumindo ter apenas constantes, e que as condições iniciais são dadas.

Supondo normalidade dos erros, temos a função de log-verossimilhança

$$\ln L(\mathbf{B}, \Sigma; \mathbf{x}_t) = -T \frac{d}{2} \ln 2\pi - T \frac{1}{2} \ln |\Sigma| - \frac{1}{2} \sum_{t=1}^T (\mathbf{x}_t - \mathbf{B}' \mathbf{Z}_t)' \Sigma^{-1} (\mathbf{x}_t - \mathbf{B}' \mathbf{Z}_t).$$

Maximizando a função de log-verossimilhança com respeito a \mathbf{B}' e Σ^{-1} , obtemos os respectivos estimadores de máxima verossimilhança:

$$\hat{\mathbf{B}} = \sum_{t=1}^T (\mathbf{Z}_t \mathbf{Z}'_t)^{-1} \sum_{t=1}^T (\mathbf{Z}_t \mathbf{x}'_t) = \mathbf{S}_{ZZ}^{-1} \mathbf{S}_{Zx}$$

$$\hat{\Sigma} = T^{-1} \sum_{t=1}^T (\mathbf{x}_t - \hat{\mathbf{B}}' \mathbf{Z}_t) (\mathbf{x}_t - \hat{\mathbf{B}}' \mathbf{Z}_t)' = \mathbf{S}_{xx} - \mathbf{S}_{xZ} \mathbf{S}_{ZZ}^{-1} \mathbf{S}_{Zx}$$

onde $\mathbf{S}_{ZZ} = \frac{1}{T} \sum_{t=1}^T (\mathbf{Z}_t \mathbf{Z}'_t)$ e $\mathbf{S}_{Zx} = \frac{1}{T} \sum_{t=1}^T (\mathbf{Z}_t \mathbf{x}'_t)$.

A seleção de p é crucial. Como no caso univariado, podemos ajustar vários modelos variando $p = 0, \dots, p_{max}$ e escolher aquele que minimiza algum critério, como por exemplo o de *Akaike* ou o de *Hannan-Quinn*:

$$AIC(p) = \log |\tilde{\Sigma}(p)| + \frac{2}{T} p \cdot d^2$$

$$HQ(p) = \log |\tilde{\Sigma}(p)| + \frac{2 \log \log T}{T} p d^2$$

onde $\tilde{\Sigma}(p) = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t \hat{\varepsilon}_t'$, a matriz de covariância dos resíduos sem a correção com os graus de liberdade.

As vezes fica difícil interpretar os diversos coeficientes do VAR(p). Podemos então fazer uma *análise estrutural* com objetivo de resumir as propriedades dinâmicas do VAR. Uma dessas análises é o teste de causalidade de Granger.

O teste de causalidade de Granger, C. W. J. (1969) é muito útil para previsões. Por exemplo, se \mathbf{x}_1 é importante para prever \mathbf{x}_2 , então dizemos que \mathbf{x}_1 Granger-Causa \mathbf{x}_2 . Mais formalmente, \mathbf{x}_1 Granger-Não-Causa \mathbf{x}_2 se para todo $s > 0$ o erro quadrático médio (EQM) da previsão de $x_{2,t+s}$ baseada em $(x_{2,t}, x_{2,t-1}, \dots)$ é o mesmo que o EQM da previsão de $x_{2,t+s}$ baseada em $(x_{2,t}, x_{2,t-1}, \dots)$ e $(x_{1,t}, x_{1,t-1}, \dots)$. Neste caso as matrizes Π_1, \dots, Π_p são triangulares. Quando \mathbf{x}_1 não Granger-Causa \mathbf{x}_2 e \mathbf{x}_2 não Granger-Causa \mathbf{x}_1 , essas matrizes serão diagonais. As hipóteses relacionadas com esses coeficientes sendo zero são testadas com o teste de Wald cuja estatística tem distribuição limite chi-quadrado, veja Tsay, R. S. (2005).

Considere o processo VAR(1)

$$\mathbf{x}_t = \mathbf{c} + \Pi \mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t \quad (6.5)$$

onde \mathbf{c} é um vetor d -dimensional, Π é uma matriz $(d \times d)$ e $\{\boldsymbol{\varepsilon}_t\}$ é uma sequência de erros não autocorrelacionados $F(\mathbf{0}, \Sigma)$. Supondo $d = 2$, o VAR(1) terá duas equações

$$x_{1,t} = c_1 + \Pi_{11}x_{1,t-1} + \Pi_{12}x_{2,t-1} + \varepsilon_{1,t}$$

$$x_{2,t} = c_2 + \Pi_{21}x_{1,t-1} + \Pi_{22}x_{2,t-1} + \varepsilon_{2,t}$$

Na primeira equação Π_{12} representa a dependência linear de $x_{1,t}$ em $x_{2,t-1}$ na presença de $x_{1,t-1}$. Assim, interpretamos Π_{12} como o efeito condicional de $x_{2,t-1}$ em $x_{1,t}$ dado $x_{1,t-1}$. Se, por exemplo na segunda equação, $\Pi_{21} = 0$, então $x_{2,t}$ não depende de $x_{1,t-1}$ quando $x_{2,t-1}$ está presente.

Olhando as duas equações conjuntamente, se, por exemplo $\Pi_{1,2} = 0$ e $\Pi_{2,1} \neq 0$ existirá uma relação unidirecional de $x_{1,t}$ para $x_{2,t}$. Se $\Pi_{2,1} = \Pi_{1,2} = 0$ então $x_{1,t}$ e $x_{2,t}$ não são linearmente dependentes. Vemos que Π mede a dependência linear dinâmica do vetor \mathbf{x}_t no seu passado. A dependência corrente, no tempo t , é medida pelos elementos fora da diagonal de Σ . É por isto que o modelo (6.5) é dito estar na sua *forma reduzida* por não mostrar explicitamente as dependências de \mathbf{x}_t no tempo t . É possível reescrever (6.5) considerando as relações no tempo t , a *forma estrutural*, mas para estimação e previsão a forma (6.5) é mais vantajosa.

6.2 Cointegração

Seja $\mathbf{x}_t = (x_{1,t}, \dots, x_{d,t})'$ um vetor $(d \times 1)$ de séries $I(1)$. O vetor \mathbf{x}_t é cointegrado se existir uma combinação linear dos componentes que seja $I(0)$, isto é, um vetor $(d \times 1)$ $\beta = (\beta_1, \dots, \beta_d)'$ tal que

$$\beta' \mathbf{x}_t = \beta_1 x_{1,t} + \dots + \beta_d x_{d,t} \sim I(0) \quad (6.6)$$

Se algum dos β_j for zero, é porque somente um subconjunto dos \mathbf{x}_t com não-zero β 's é cointegrado.

A combinação linear $\beta' \mathbf{x}_t$ em geral tem interpretação econômica e é chamada de *relação de equilíbrio de longo prazo*. A interpretação é que, no longo prazo, existirão forças econômicas forçando as variáveis $I(1)$ a retornarem à(s) sua(s) relação(ões) de equilíbrio.

β é chamado de vetor cointegrante e não é único já que para qualquer constante c temos $\beta^* = c\beta$ e $\beta^* \mathbf{x}_t \sim I(0)$. Uma normalização usual para unicamente identificar β é:

$$\beta = (1, -\beta_2, \dots, -\beta_d)'$$

Assim, (6.6) fica

$$\beta' \mathbf{x}_t = x_{1,t} - \beta_2 x_{2,t} - \dots - \beta_d x_{d,t} \sim I(0) \quad (6.7)$$

ou, interpretando o lado direito de (6.7) como os resíduos de uma regressão:

$$x_{1,t} = \beta_2 x_{2,t} + \dots + \beta_d x_{d,t} + u_t \quad (6.8)$$

onde $u_t \sim I(0)$.

Em (6.8) o erro u_t é chamado erro de disequilíbrio (*disequilibrium error*) ou ainda de resíduo de cointegração (*cointegrating residual*). No longo prazo o erro de disequilíbrio u_t é zero e a *relação de equilíbrio de longo prazo* é

$$x_{1,t} = \beta_2 x_{2,t} + \dots + \beta_d x_{d,t}.$$

Em geral, para $d > 2$ temos várias relações de cointegração e não apenas uma. Neste caso, se o vetor $(d \times 1)$ \mathbf{x}_t é cointegrado, podem existir r linearmente independentes vetores cointegrantes, $0 < r < d$. Por exemplo (veja ilustração no final deste capítulo), se $d = 3$ e $r = 2$ teremos 2 vetores cointegrantes

$$\beta_1 = (\beta_{11}, \beta_{12}, \beta_{13})'$$

$$\beta_2 = (\beta_{21}, \beta_{22}, \beta_{23})',$$

tais que $\beta_1' \mathbf{x}_t \sim I(0)$ e $\beta_2' \mathbf{x}_t \sim I(0)$. Neste caso a matriz (3×2) \mathbf{B}

$$\mathbf{B}' = \begin{pmatrix} \beta_1' \\ \beta_2' \end{pmatrix} = \begin{pmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \end{pmatrix}$$

forma uma base para os vetores cointegrantes, que não é única (assim como β_1 e β_2 não são) e normalizações são necessárias.

É interessante ver o problema dual: se o vetor $(d \times 1)$ \mathbf{x}_t é cointegrado com $0 < r < d$ vetores cointegrantes, então existem $(d - r)$ *tendências estocásticas comuns que são* $I(1)$. No exemplo acima

com $d = 3$ e $r = 2$, teremos 1 tendência estocástica comum às três séries. Existe uma representação devida à Phillips, P. C. B. (1991), denominada representação *triangular*, a qual no exemplo, impõe que $\beta_1 = (1, 0, -\beta_{13})'$, $\beta_2 = (0, 1, -\beta_{23})'$ e portanto

$$x_{1,t} = \beta_{13}x_{3,t} + u_t, \quad u_t \sim I(0)$$

$$x_{2,t} = \beta_{23}x_{3,t} + v_t, \quad v_t \sim I(0)$$

$$x_{3,t} = x_{3,t-1} + w_t, \quad w_t \sim I(0),$$

onde as duas primeiras equações descrevem as duas relações de equilíbrio de longo prazo, e a terceira fornece a tendência estocástica comum. Um exemplo dado em Zivot, E. e Wang, J. (2006) de um sistema deste tipo consiste em um sistema da estrutura a termo de taxas de juros, onde x_3 representa a taxa de curto prazo, e x_1 e x_2 representam duas taxas diferentes de longo prazo.

6.2.1 O modelo VECM

Considere o modelo VAR(p) dado em (6.1) e (6.3), com a generalização para a parte determinística

$$\mathbf{x}_t = \Phi \mathbf{D}_t + \Pi_1 \mathbf{x}_{t-1} + \dots + \Pi_p \mathbf{x}_{t-p} + \varepsilon_t, \quad t = 1, \dots, T, \quad (6.9)$$

onde \mathbf{D}_t contém termos determinísticos tais como constante, tendência, *dummies* sazonais, etc. Lembremos que o modelo é estável se as raízes de $\det(\mathbf{I}_d - \Pi_1 z - \dots - \Pi_p z^p) = 0$ estão todas fora do círculo complexo unitário. Se alguma das raízes estiver sobre o círculo, todas ou algumas das variáveis \mathbf{x}_t serão $I(1)$ e podem ser cointegradas. Neste caso a representação VAR(p) de (6.9) não é adequada pois existem as relações de cointegração e devemos considerar o *vector error correction model*, modelo de correção do erro de equilíbrio (Engle, R. F. e Granger, C. W. J. (1987)), VECM($p - 1$).

$$\Delta \mathbf{x}_t = \Phi \mathbf{D}_t + \Pi \mathbf{x}_{t-1} + \Gamma_1 \Delta \mathbf{x}_{t-1} + \dots + \Gamma_{p-1} \Delta \mathbf{x}_{t-p+1} + \varepsilon_t \quad (6.10)$$

onde

$$\Pi = \Pi_1 + \dots + \Pi_p - \mathbf{I}_d$$

e

$$\Gamma_k = - \sum_{j=k+1}^p \Pi_j, \quad k = 1, \dots, p-1.$$

Notemos que a equação (6.10) pode ser modificada de várias formas diferentes, inclusive com a introdução de regressores determinísticos, com a adição de termos defasados em $\Delta \mathbf{x}_t$, ou ainda permitindo que as componentes de \mathbf{x}_t sejam integradas de ordem maior que 1.

A matriz Π é a *matriz de impacto de longo prazo*, e as matrizes Γ_k são as matrizes de *impacto de curto prazo*. Esta especificação corresponde à especificação “*transitory*” do pacote R. Notemos que os parâmetros Π_i do VAR podem ser obtidos do VECM através dos parâmetros Π e Γ_k :

$$\Pi_1 = \Gamma_1 + \Pi + \mathbf{I}_d$$

$$\Pi_k = \Gamma_k - \Gamma_{k-1}, \quad k = 2, \dots, p.$$

No modelo (6.10), $\Delta \mathbf{x}_t$ e os seus valores defasados são $I(0)$; \mathbf{x}_{t-1} são variáveis $I(1)$ e o termo $\Pi \mathbf{x}_{t-1}$ é $I(0)$ e contém as relações de cointegração. Se a matriz Π for singular, terá posto reduzido, isto é, $\text{posto}(\Pi) = r < d$. Temos dois casos a considerar: (i) $\text{posto}(\Pi) = 0$. Isto implica que $\Pi = 0$ e assim $\mathbf{x}_t \sim I(1)$ mas *não* cointegradas. O VECM (6.10) então se torna um VAR($p-1$) para as primeiras diferenças:

$$\Delta \mathbf{x}_{t-1} = \Phi \mathbf{D}_t + \Gamma_1 \Delta \mathbf{x}_{t-1} + \dots + \Gamma_{p-1} \Delta \mathbf{x}_{t-p+1} + \varepsilon_t.$$

(ii) $0 < \text{posto}(\Pi) = r < d$. Neste caso $\mathbf{x}_t \sim I(1)$ com r vetores de cointegração linearmente independentes, e $d-r$ tendências estocásticas comuns. Podemos escrever Π como

$$\Pi_{(d \times d)} = \alpha_{(d \times r)} \beta'_{(r \times d)}$$

onde α e β são matrizes $(d \times r)$ com $\text{posto}(\alpha) = \text{posto}(\beta) = r$. As linhas de β' formam uma base para os r vetores cointegrantes, e os elementos de α distribuem o impacto dos vetores de cointegração na evolução de $\Delta \mathbf{x}_t$ e são interpretados como *coeficientes da velocidade de ajustamento*. O modelo VECM (6.10) se torna:

$$\Delta \mathbf{x}_t = \Phi \mathbf{D}_t + \alpha \beta' \mathbf{x}_{t-1} + \Gamma_1 \Delta \mathbf{x}_{t-1} + \dots + \Gamma_{p-1} \Delta \mathbf{x}_{t-p+1} + \varepsilon_t \quad (6.11)$$

com $\beta' \mathbf{x}_{t-1} \sim I(0)$. A fatorização $\Pi = \alpha \beta$ não é única. Para obter valores únicos para α e β é preciso impor restrições ao modelo.

Considere, por exemplo, $d = 2$, $p = 1$. Neste caso o modelo VAR(1) cointegrado para $(x_{1,t}, x_{2,t})$ seria

$$\mathbf{x}_t = \Pi_1 \mathbf{x}_{t-1} + \varepsilon_t,$$

e o VECM($p-1$) é

$$\Delta \mathbf{x}_t = \Pi \mathbf{x}_{t-1} + \varepsilon_t,$$

onde $\Pi = \Pi_1 - \mathbf{I}_2$. Sendo \mathbf{x}_t cointegrado, temos um vetor cointegrante β tal que $\beta_1 x_{1,t} + \beta_2 x_{2,t} \sim I(0)$. Usando a normalização $\beta = (1, -\beta)$ obtemos a relação estocástica de equilíbrio de longo prazo

$$x_{1,t} = \beta x_{2,t} + u_t$$

onde $u_t \sim I(0)$ e u_t representa os desvios estocásticos do equilíbrio de longo prazo $x_{1,t} = \beta x_{2,t}$.

A matriz Π é (2×2) e como temos um vetor de cointegração, $\text{posto}(\Pi) = 1$ e obtemos a decomposição

$$\Pi = \alpha \beta' = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} (1, -\beta) = \begin{bmatrix} \alpha_1 & -\alpha_1 \beta \\ \alpha_2 & -\alpha_2 \beta \end{bmatrix}$$

e

$$\Pi \mathbf{x}_{t-1} = \begin{bmatrix} \alpha_1 & -\alpha_1 \beta \\ \alpha_2 & -\alpha_2 \beta \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} = \begin{bmatrix} \alpha_1 x_{1,t-1} & -\alpha_1 \beta x_{2,t-1} \\ \alpha_2 x_{1,t-1} & -\alpha_2 \beta x_{2,t-1} \end{bmatrix}$$

e podemos reescrever o VECM como

$$\Delta \mathbf{x}_t = \alpha \beta' \mathbf{x}_{t-1} + \varepsilon_t \quad (6.12)$$

ou dar as duas equações

$$\Delta x_{1,t} = \alpha_1(x_{1,t-1} - \beta x_{2,t-1}) + \varepsilon_{1,t}$$

$$\Delta x_{2,t} = \alpha_2(x_{1,t-1} - \beta x_{2,t-1}) + \varepsilon_{2,t}$$

onde $(x_{1,t-1} - \beta x_{2,t-1})$ é o erro de desequilíbrio defasado, comum às duas equações. O importante é que as reações de \mathbf{x}_1 e de \mathbf{x}_2 aos erros de desequilíbrio são capturadas pelos coeficientes de ajustamento α_1 e α_2 . Sendo α uma matriz $(d \times r)$, definimos α_\perp uma matriz $d \times (d - r)$ tal que $\alpha \alpha_\perp = 0$. Convencionamos $\alpha_\perp = I$ se $\alpha = 0$ e se $\alpha = I$ então $\alpha_\perp = 0$. As tendências comuns são $\alpha_\perp \sum_{i=1}^T \varepsilon_i$.

A interpretação é que a representação das soluções do modelo autorregressivo VECM em termos das tendências comuns, mostra que a não estacionariedade nas variáveis é criada pelos choques não antecipados e acumulados do processo, mas que nem todos os choques aparecem. Eles aparecem multiplicados por uma matriz α_\perp a qual mostra que apenas $(d - r)$ passeios aleatórios dão origem à não estacionariedade. As combinações lineares $\beta \mathbf{x}_t$ não são influenciadas pelos passeios aleatórios e se tornam estacionárias.

As equações acima que descrevem o comportamento dinâmico de $\Delta x_{1,t}$ e $\Delta x_{2,t}$ podem ser aumentadas com a inclusão de termos determinísticos e de defasagens. Note que o modelo VECM liga a relação de equilíbrio de longo prazo que existe devida à cointegração com o mecanismo dinâmico de ajustamento de curto prazo α_1 e α_2 . Os coeficientes de ajustamento α_1 e α_2 descrevem como as variáveis reagem quando elas se afastam do equilíbrio de longo prazo, ou suas reações ao erro de desequilíbrio.

Ambas contém o termo $(x_{1,t-1} - \beta x_{2,t-1})$ que é o erro de desequilíbrio defasado. Os α_i 's controlam a velocidade com a qual os $x_{i,t}$ respondem ao erro de desequilíbrio. Quando as variáveis estão fora do seu equilíbrio de longo prazo, as forças econométricas representadas pelos α 's empurram o modelo de volta ao seu equilíbrio de longo prazo. Por exemplo se $\alpha_i = 0.5$, aproximadamente metade do erro de desequilíbrio é corrigido em um período. Se $\alpha_i \sim 1$, aproximadamente corrigido por completo. Se $\alpha_i > 1$ a correção supera à necessidade de correção.

Uma propriedade importante é que cointegração é invariante às extensões do conjunto de informações utilizado, isto é, se mais variáveis são incluídas na análise, os vetores de cointegração ainda estarão presentes, mas as tendências comuns podem mudar completamente, já que temos mais fatores explicativos.

Em resumo, o modelo VECM na sua forma reduzida deve especificar que $\Pi = \alpha\beta$ na sua formulação, e supor que as raízes das equações estejam fora do disco unitário, e que os parâmetros $\alpha, \beta, \Gamma_1, \dots, \Gamma_k, \Phi$ e Σ variam livremente, podendo ser estimados. A suposição de normalidade permite escrever a função de verossimilhança. Notemos porém, que os parâmetros α e β são não identificáveis. Assim, a inferência é feita nas equações de cointegração, $sp(\beta)$, e no de ajustamento, $sp(\alpha)$, e é preciso normalizar β .

Seja $0 < rank(\Pi) = r < d$. Os estimadores de máxima verossimilhança $\hat{\beta}_{EMV}$ são os r autovetores associados aos maiores auto-valores $\hat{\lambda}_1 > \dots > \hat{\lambda}_r$. Tendo estimado os $\hat{\beta}_{EMV}$ os parâmetros restantes são obtidos por mínimos quadrados. Como a fatorização $\Pi = \alpha\beta'$ não é única, as colunas de $\hat{\beta}_{EMV}$ são interpretadas como combinações lineares das verdadeiras relações de cointegração. Para interpretações é sempre conveniente normalizar esses vetores de acordo com algum sistema de coordenadas. Uma

dessas sugestões é achar a representação triangular do sistema cointegrado, o que dá origem ao $\widehat{\beta}_{c,EMV}$. Esta normalização afeta os EMV de α .

Notemos que, uma vez que as r relações de cointegração tenham sido achadas, podemos estabelecer e testar diversas hipóteses sobre os coeficientes α e β , veja Johansen, S. (1995). O modelo VECM com $\Pi = \alpha\beta$ é uma regressão de posto reduzido cuja solução pode ser obtida através de seus autovalores. Veja detalhes em Johansen, S. e Juselius, K. (1990). Num primeiro passo, os parâmetros autorregressivos Γ_i e determinísticos Φ são estimados. Em seguida, os resíduos são utilizados para obter os restantes. A metodologia permite testar os modelos encaixados. As distribuições assintóticas dos estimadores são dadas em Johansen, S. (1995). A partir de suposições simples, a distribuição assintótica de $\widehat{\beta}$ é aproximadamente Normal.

6.2.2 Testes para cointegração

Considere novamente $\mathbf{x}_t \in \mathbb{R}^d$, $\mathbf{x}_t \sim I(1)$ e $0 < r < d$, e assim temos r vetores cointegrantes. Seja \mathbf{B}' uma matriz ($r \times d$) contendo os vetores cointegrantes. Então:

$$\mathbf{B}'\mathbf{x}_t = \begin{pmatrix} \beta'_1 & \mathbf{x}_t \\ \beta'_2 & \mathbf{x}_t \\ \vdots & \\ \beta'_r & \mathbf{x}_t \end{pmatrix} = \begin{pmatrix} u_{1,t} \\ u_{2,t} \\ \vdots \\ u_{r,t} \end{pmatrix} \sim I(0)$$

Os testes mais importantes para se testar se existe cointegração, ou se existe alguma(s) combinação(ões) linear(es) $I(0)$ dos componentes de \mathbf{x}_t , são os de Engle, R. F. e Granger, C. W. J. (1987) (esta referência daqui para frente será denotada por EG87) e de Johansen, S. (1988).

O teste em dois passos de EG87 assume que exista no máximo um vetor cointegrante. Os dois passos são (i) Ache os resíduos cointegrantes $\widehat{u}_t = \widehat{\beta}'\mathbf{x}_t$. (ii) Teste se u_t é $I(0)$ (teste da raiz unitária, RU).

A hipótese nula do teste é H_0 : Não há cointegração ou $u_t \sim I(1)$ e H_1 : Há cointegração ou $u_t \sim I(0)$. Dois casos podem ser considerados. No primeiro β é conhecido ou especificado. O segundo estima β . Se β for especificado, use-o para obter os resíduos do passo 1. Qualquer teste para detectar RU pode ser usado no passo 2. Em geral usa-se o teste ADF ou o PP, vistos no Capítulo 4. É muito importante especificar corretamente a parte determinística desses testes, assim como o número de defasagens das variáveis.

No caso de β ser desconhecido deve ser estimado, mas antes algum tipo de padronização deve ser feita para se ter unicidade na estimação. Em geral, coloca-se x_1 como variável dependente e (x_2, x_3, \dots, x_d) como variáveis explanatórias. Então β é normalizado como $\beta = (1, -\beta'_2)'$ e temos a regressão

$$x_{1,t} = c + (\beta_2 x_{2,t} + \beta_3 x_{3,t} + \dots + \beta_d x_{d,t}) + u_t \quad (6.13)$$

que pode ser estimada por MQO. Os resíduos

$$\widehat{u}_t = x_{1,t} - \widehat{c} - (\widehat{\beta}_2 x_{2,t} + \dots + \widehat{\beta}_d x_{d,t}) \quad (6.14)$$

são usados em um teste de RU para detectar se são $I(0)$. O teste de RU neste caso deve ser especificado sem termos determinísticos, constante e/ou tendência, sempre que a média de $\{\hat{u}_t\}$ for zero. Notando que sob H_0 teríamos um caso de regressão espúria, Phillips, P. C. B. e Ouliaris, S. (1990) mostraram que os testes ADF e PP aplicados aos resíduos (6.14) não seguiam a distribuição usual de Dickey-Fuller (sob H_0). A distribuição (PO) seria uma função do processo de Wiener e dependente da especificação do termo determinístico em (6.13), assim como do número $(d-1)$ de variáveis, veja Phillips, P. C. B. e Ouliaris, S. (1990) e Hansen, B. (1994). Para a definição final da distribuição, três possibilidades para as tendências de \mathbf{x}_t devem ser consideradas:

Caso 1. Todas as d variáveis são $I(1)$ sem *drift*. Neste caso um vetor $(d-1) \times 1$ de constantes deve ser considerado na distribuição PO da estatística teste.

Caso 2. Todas as $(d-1)$ variáveis (x_2, \dots, x_d) são $I(1)$ com *drift*, e x_1 pode ou não ter *drift*. Neste caso a distribuição PO da estatística teste sob H_0 deve considerar uma constante e tendência de dimensão $d-2$.

Caso 3. $x_{1,t}$ tem *drift* e $(x_{2,t}, \dots, x_{d,t})$ e não possuem. Neste caso β_2 deve ser estimado da regressão

$$x_{1,t} = c + \delta t + \beta'_2(x_{2,t}, \dots, x_{d,t}) + u_t.$$

Neste caso, a distribuição PO da estatística teste de ADF ou PP deve ser ajustada para uma constante e uma tendência, de dimensão $d-1$. O SPlus eo R consideram todos esses casos fornecendo os valores críticos da distribuição.

Sobre as propriedades assintóticas dos estimadores de MQ ao serem usados para estimar os vetores cointegrantes veja Stock, J. H. (1987). Em resumo a distribuição assintótica de $T(\hat{\beta}_2 - \beta_2)$ não é a Normal e a fórmula usual para calcular a variância assintótica não é válida.

Um estimador com melhores propriedades considera termos determinísticos e termos defasados e adiantados na regressão (veja em Hendry, D. F. e Juselius, K. (2000)). A regressão aumentada seria

$$x_{1,t} = \gamma' \mathbf{D}_t + (\beta_2, \dots, \beta_d)'(x_{2,t}, \dots, x_{d,t}) + \sum_{j=-p}^p \psi'_j \Delta(x_{2,t}, \dots, x_{d,t}) + u_t \quad (6.15)$$

a qual é estimada por MQ. Este é o MQ dinâmico (DOLS), eficiente e assintoticamente Normal.

O teste de Johansen

Seja o modelo VECM (6.10) e chamemos este modelo de $H(r)$. Temos então os modelos encaixados

$$H(0) \subset \dots \subset H(r) \subset \dots \subset H(d)$$

onde $H(0)$ representa o VAR não cointegrado com $\mathbf{\Pi} = 0$, e $H(d)$ representa o modelo VAR(p) estacionário não-restrito. A metodologia de Johansen é baseada nesta sequência. Como o posto da matriz $\mathbf{\Pi}$ de impacto de longo prazo dá o número de relações de cointegração, o teste de Johansen baseia-se nos autovalores $\hat{\lambda}_1 > \hat{\lambda}_2 > \dots > \hat{\lambda}_d$ da matriz $\mathbf{\Pi}$, os quais estão entre 0 e 1, e tais que $\text{posto}(\mathbf{\Pi})$ é igual ao número de autovalores não nulos.

Em resumo, a metodologia de Johansen consiste em: (i) Especificar e estimar um modelo VAR(p) para \mathbf{x}_t , equação (6.9). (ii) Determinar o posto r da matriz $\mathbf{\Pi}$ em (6.10). Para isto, estima-se

por máxima verossimilhança um sistema de duas equações de regressão que considera as defasagens $\Delta \mathbf{x}_{t-j}$, $j = 1, 2, \dots, p-1$, para explicar $\Delta \mathbf{x}_t$ e \mathbf{x}_t . Obtém-se as matrizes de covariâncias amostrais baseadas nas duas séries de resíduos, recaindo-se em um problema de determinar os autovetores de uma matriz. O número de autovalores diferentes de zero é o valor de r . Os r autovetores são os r vetores de cointegração não normalizados. (iii) Identificar restrições e impor normalizações para os vetores cointegrantes. (iv) Utilizando os vetores de cointegração normalizados, estimar por máxima verossimilhança os parâmetros restantes do VECM correspondente. (v) O valor máximo da função de verossimilhança baseada nos r vetores cointegrantes é usado na construção de testes estatísticos. A aplicação do teste é sequencial com a hipótese nula variando desde $r = 0$ até $r = r_0$. A não rejeição da hipótese nula define o número final de relações de cointegração.

O teste de Johansen tem hipóteses

$$H_0 : r = r_0, \quad H_1 : r > r_0.$$

A estatística do traço, *trace statistic*, é dada por

$$LR_{trace}(r_0) = -T \sum_{i=r_0+1}^d \log(1 - \hat{\lambda}_i).$$

Se posto($\mathbf{\Pi}$) = r_0 , então $\hat{\lambda}_{r_0+1}, \dots, \hat{\lambda}_d$ devem ser todos quase zero e a estatística $LR_{trace}(r_0)$ deve ser pequena. A distribuição assintótica da estatística sob H_0 não é chi-quadrado (veja Tsay, R. S. (2005)), e depende de $(d - r_0)$ e da especificação do termo determinístico. Valores críticos estão tabelados em Osterwald-Lenum(1992) para os cinco casos previstos (veja abaixo). O teste se inicia supondo $r_0 = 0$ e continua sequencialmente até que H_0 seja aceita.

A outra estatística teste de Johansen é do máximo autovalor:

$$LR_{max}(r_0) = -T \log(1 - \hat{\lambda}_{r_0+1}),$$

e testa

$$H_0 : r = r_0 \quad H_1 : r = r_0 + 1,$$

e novamente os valores críticos para os cinco casos previstos estão tabelados.

De acordo com a metodologia de Johansen, os termos determinísticos em (6.11) são restritos a serem desta forma:

$$\Phi \mathbf{D}_t = \boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t.$$

Se os termos determinísticos forem não-restritos, então as séries \mathbf{x}_t poderão exibir tendências quadráticas e mais, poderemos ter um termo de tendência linear nas relações cointegrantes. Colocando restrições em $\boldsymbol{\mu}_0$ e $\boldsymbol{\mu}_1$ limitaremos as tendências de \mathbf{x}_t . Assim, podemos classificar o comportamento das tendências de \mathbf{x}_t em cinco casos:

(1) Modelo $H_2(r) : \boldsymbol{\mu}_t = 0$ (sem constante). Neste caso o VECM restrito é

$$\Delta \mathbf{x}_t = \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{x}_{t-1} + \boldsymbol{\Gamma}_1 \Delta \mathbf{x}_{t-1} + \dots + \boldsymbol{\Gamma}_{p-1} \Delta \mathbf{x}_{t-p+1} + \boldsymbol{\varepsilon}_t$$

e todas as séries em \mathbf{x}_t são I(1) *sem drift* e as relações $\beta' \mathbf{x}_t$ têm média zero. (2) Modelo $H_1^*(r) : \mu_t = \mu_0 = \alpha \rho_0$ (constante restrita). O VECM restrito é:

$$\Delta \mathbf{x}_t = \alpha(\beta' \mathbf{x}_{t-1} + \rho_0) + \Gamma_1 \Delta \mathbf{x}_{t-1} + \dots + \Gamma_{p-1} \Delta \mathbf{x}_{t-p+1} + \varepsilon_t$$

isto é, incluímos a constante no vetor cointegrante. As séries \mathbf{x}_t são I(1) *sem drift* e as relações cointegrantes $\beta' \mathbf{x}_t$ terão médias ρ_0 não zero (ρ_0 é $(r \times 1)$).

(3) Modelo $H_1(r) : \mu_t = \mu_0$ (constante não-restrita). O VECM restrito é

$$\Delta \mathbf{x}_t = \mu_0 + \alpha \beta' \mathbf{x}_{t-1} + \Gamma_1 \Delta \mathbf{x}_{t-1} + \dots + \Gamma_{p-1} \Delta \mathbf{x}_{t-p+1} + \varepsilon_t$$

onde $\mathbf{x}_t \sim I(1)$ com vetor de *drifts* μ_0 (μ_0 é $(d \times 1)$) e as relações $\beta' \mathbf{x}_t$ podem ter média não-zero.

(4) Modelo $H^*(r) : \mu_t = \mu_0 + \alpha \rho_1 t$ (tendência restrita). O VECM restrito será

$$\Delta \mathbf{x}_t = \mu_0 + \alpha(\beta' \mathbf{x}_{t-1} + \rho_1 t) + \Gamma_1 \Delta \mathbf{x}_{t-1} + \dots + \Gamma_{p-1} \Delta \mathbf{x}_{t-p+1} + \varepsilon_t$$

onde $\mathbf{x}_t \sim I(1)$ com vetor de *drifts* μ_0 e as relações de cointegração $\beta' \mathbf{x}_t$ terão um termo de tendência linear $\rho_1 t$, ρ_1 é $(r \times 1)$.

(5) Modelo $H(r) : \mu_t = \mu_0 + \mu_1 t$ (tendência e constante não-restritas). O VECM não-restrito é

$$\Delta \mathbf{x}_t = \mu_0 + \mu_1 t + \alpha \beta' \mathbf{x}_{t-1} + \Gamma_1 \Delta \mathbf{x}_{t-1} + \dots + \Gamma_{p-1} \Delta \mathbf{x}_{t-p+1} + \varepsilon_t$$

onde $\mathbf{x}_t \sim I(1)$ com tendência linear (quadrática em \mathbf{x}_t) e $\beta' \mathbf{x}_t$ tem tendência linear.

O caso (1) é apropriado para quando as séries \mathbf{x}_t não apresentam tendência. Caso (2) é apropriado para $\mathbf{x}_t \sim I(1)$ sem tendência. Caso (3) é apropriado para $\mathbf{x}_t \sim I(1)$ como por exemplo preço de ações, agregados macroeconômicos. Caso (5) é apropriado para séries $\mathbf{x}_t \sim I(1)$ com tendência quadrática.

6.2.3 Regressão espúria

Quando todas as variáveis em um modelo de regressão são I(0), os resultados estatísticos para o modelo linear de regressão tem propriedades bem conhecidas. Quando todas são I(1), porém *não* cointegradas, temos o fenômeno da regressão espúria.

Nesta situação temos a regressão através de MQ fornecendo resultados considerados bons, tais como estatísticas t significativas, R^2 interessante. Porém, a estatística de Durbin-Watson pode ser baixa, sugerindo forte autocorrelação nos resíduos. A análise dos resíduos (que sempre deve ser efetuada) deve incluir o exame da f.a.c., a qual, neste caso, irá mostrar vários coeficientes de autocorrelação significativos. Neste caso, a abordagem correta é utilizar as diferenças das variáveis e o modelo VARMA. Boas referências são Hamilton, J. D. (1994) e Phillips, P. C. B. (1986).

6.3 Cointegração com limiar

Nesta seção revemos alguns conceitos já dados nas duas seções anteriores. O motivo é fazer com que o leitor encontre aqui as informações na perspectiva necessária de modo a definir um modelo com limiar, não linear, para séries cointegradas.

Suponha um exemplo simples onde duas séries $\{x_{1,t}\}$ e $\{x_{2,t}\}$ possuam raiz unitária, e portanto são não estacionárias. Isto é, ambas seguem algum $\text{ARIMA}(p, 1, q)$. Contudo, se o sistema formado pelas duas séries possui apenas uma raiz unitária, elas são ditas cointegradas. Neste caso, existe uma combinação linear das duas séries que é $I(0)$, isto é, estacionária de raiz unitária.

Generalizando, d séries não estacionárias de raiz unitária são ditas cointegradas se existirem menos que d raízes unitárias no sistema, digamos $0 < r < d$. Neste caso não adianta tomar as primeiras diferenças e considerar um VAR pois estaremos super-diferenciando. A quantidade $d - r$ é chamada de número de fatores cointegrantes. Este número r é o número de combinações lineares diferentes que são $I(0)$.

Assim, se $\mathbf{x}_t = (x_{1,t}, \dots, x_{d,t})'$ é um vetor $(d \times 1)$ de séries $I(1)$, \mathbf{x}_t é cointegrado se existir um vetor $(d \times 1)$ $\boldsymbol{\beta} = (\beta_1, \dots, \beta_d)$ tal que

$$\boldsymbol{\beta}'\mathbf{x}_t = \beta_1 x_{1,t} + \dots + \beta_d x_{d,t} \sim I(0). \quad (6.16)$$

Se alguns dos β_j for zero, é porque apenas um subconjunto das $x_{j,t}$ são cointegradas. $\boldsymbol{\beta}'\mathbf{x}_t$ representa a relação de equilíbrio no longo prazo.

Como (6.16) não é único, requer uma normalização. Em geral, se faz $\boldsymbol{\beta} = (1, -\beta_2, \dots, -\beta_d)'$. Assim, (6.16) fica

$$\boldsymbol{\beta}\mathbf{x}_t = x_{1,t} - \beta_2 x_{2,t} - \dots - \beta_d x_{d,t} \sim I(0)$$

e podemos escrever

$$x_{1,t} = \beta_2 x_{2,t} + \dots + \beta_d x_{d,t} + u_t \quad (6.17)$$

onde $u_t \sim I(0)$. O erro u_t é chamado o *erro de desequilíbrio* ou de *resíduo da cointegração*.

Generalizando, podem existir $0 < r < d$ vetores de cointegração linearmente independentes. Representamos como

$$\mathbf{B}' = \begin{pmatrix} \boldsymbol{\beta}'_1 \\ \vdots \\ \boldsymbol{\beta}'_r \end{pmatrix} = \begin{pmatrix} \beta_{1,1} & \beta_{1,2} & \cdots & \beta_{1,d} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{r,1} & \beta_{r,2} & \cdots & \beta_{r,d} \end{pmatrix}$$

Novamente não são únicos e, além disto, qualquer combinação linear de $\boldsymbol{\beta}_i$ e $\boldsymbol{\beta}_j$ é também vetor de cointegração.

O modelo clássico (visto na Seção 6.2) para variáveis cointegradas assume linearidade e ajustamento simétrico. Em sua forma simples, o modelo pode ser escrito como

$$\Delta\mathbf{x}_t = \boldsymbol{\Pi}\mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t \quad (6.18)$$

onde \mathbf{x}_t é um vetor $(d \times 1)$ de variáveis aleatórias, todas integradas de ordem 1, $\boldsymbol{\Pi}$ é uma matriz $(d \times d)$, e $\boldsymbol{\varepsilon}_t$ é um vetor $(d \times 1)$ de erros aleatórios podendo ser correlacionados no tempo. Conforme vimos, o teste de Johansen baseia-se na estimação $\boldsymbol{\Pi}$ e na determinação do seu posto.

A equação (6.18) pode ser estendida de várias formas, introduzindo regressores determinísticos, incluindo valores defasados de $\Delta\mathbf{x}_t$, e permitindo que os componentes do vetor \mathbf{x}_t sejam integrados de ordem maiores que 1. A suposição clássica implícita em (6.18) é que, se $\text{posto}(\boldsymbol{\Pi}) \neq 0$, então o sistema exhibe ajustamento simétrico em torno de $\mathbf{x}_t = 0$, e que para qualquer $\mathbf{x}_t \neq 0$, $\Delta\mathbf{x}_{t+1}$ sempre será igual a $\boldsymbol{\Pi}\mathbf{x}_t$.

A alternativa do teste de EG87, também assume ajustamento simétrico. Na sua forma mais simples, o teste em dois passos de EG87 estima por mínimos quadrados ordinários (MQO), no primeiro passo, a relação de equilíbrio de longo prazo (estamos supondo que $r = 1$)

$$x_{1,t} = \beta_0 + \beta_2 x_{2,t} + \dots + \beta_d x_{d,t} + u_t \quad (6.19)$$

onde x_{it} são as componentes I(1) do vetor \mathbf{x}_t , β_i são parâmetros e u_t é o erro aleatório que pode ser correlacionado no tempo.

No segundo passo, estima-se por MQO o parâmetro ρ da equação de regressão:

$$\Delta u_t = \rho u_{t-1} + e_t \quad (6.20)$$

onde e_t é um ruído branco. No processo de estimação de (6.20) são usados os resíduos de (6.19). As hipóteses do teste são

$$H_0 : \text{Não há cointegração} \quad H_1 : -2 < \rho < 0.$$

Rejeitar H_0 significa que os resíduos de (6.19) são estacionários e possuem média zero. Neste caso (6.19) é um atrator, cuja força de atração é proporcional ao valor absoluto de u_t . O teorema da representação de Granger garante que se $\rho \neq 0$, as equações (6.19) e (6.20) conjuntamente implicam na existência de um modelo de correção do erro (ECM) dado por

$$\Delta x_{i,t} = \alpha_i (x_{1,t-1} - \beta_0 - \beta_2 x_{2,t-1} - \dots - \beta_d x_{d,t-1}) + \dots + \varepsilon_{i,t}. \quad (6.21)$$

Enders, W. e Siklos, P. L. (2001) comentam que o teste de EG87 não se adequa às situações onde o ajustamento é assimétrico. Eles propõem um outro teste baseado numa especificação alternativa do ECM, chamado de TAR (*threshold autoregressive*), onde a equação (6.20) é reescrita como

$$\Delta u_t = I_t \rho_1 u_{t-1} + (1 - I_t) \rho_2 u_{t-1} + e_t \quad (6.22)$$

onde I_t é a função indicadora

$$I_t = \begin{cases} 1, & \text{se } u_{t-1} \geq \tau \\ 0, & \text{se } u_{t-1} < \tau \end{cases}, \quad (6.23)$$

onde τ é o valor de um limiar e $\{e_t\}$ é uma sequência de v.a.'s i.i.d. $(0, \sigma_e^2)$, sendo e_t independente de u_j para $j < t$.

Um resultado importante é o de Petrucelli, J. e Woolford, S. (1984), que diz que as condições necessárias e suficientes para a estacionariedade de $\{u_t\}$ são $\rho_1 < 0$, $\rho_2 < 0$, e $(1 + \rho_1)(1 + \rho_2) < 1$, para qualquer valor de τ . Se as condições se verificam, temos que $u_t = 0$ representa o valor de equilíbrio do sistema no longo prazo quando $x_{1,t} = \beta_0 + \beta_2 x_{2,t} + \dots + \beta_d x_{d,t}$. Notemos que se $\rho_1 = \rho_2$ temos recuperado o teste original de EG87. A distribuição assintótica dos estimadores de MQ de ρ_1 e ρ_2 é a Normal multivariada. Em inúmeras aplicações econômicas faz sentido ter-se $\tau = 0$. Neste caso, o vetor de cointegração coincide com o atrator. O ajustamento é $\rho_1 u_{t-1}$ se u_{t-1} está acima do valor do equilíbrio de longo prazo, e será $\rho_2 u_{t-1}$ se abaixo.

Supondo que exista um único vetor de cointegração $(\beta_0, \beta_2, \dots, \beta_d)$, o modelo VECM para cada variável $x_{i,t}$ pode ser escrito como:

$$\Delta x_{i,t} = \rho_{1,i} I_t u_{t-1} + \rho_{2,i} (1 - I_t) u_{t-1} + \dots + \varepsilon_{i,t} \quad (6.24)$$

onde $\rho_{1,i}$ e $\rho_{2,i}$ representam a velocidade de ajustamento de $\Delta x_{i,t}$, coeficientes estes que podem ser diferentes para cada variável i . Para achar (6.24) usamos (6.19), isto é, $u_{t-1} = (x_{1,t-1} - \beta_0 - \beta_2 x_{2,t-1} - \dots - \beta_d x_{d,t-1})$. A Figura 6.1 mostra as trajetórias no tempo de duas séries I(1) cointegradas com limiar (modelo TAR).

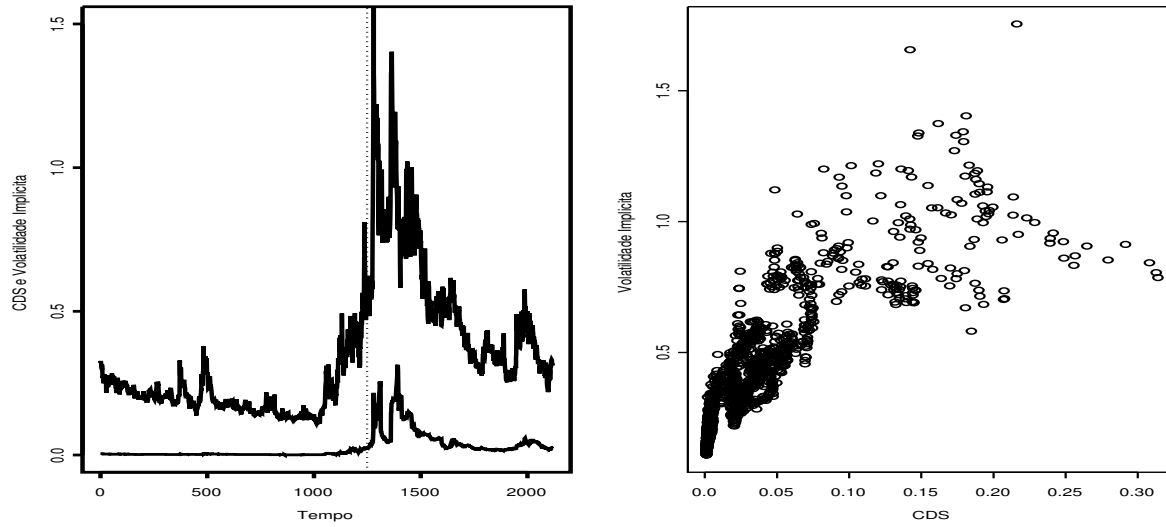


Figura 6.1: Duas séries cointegradas. Séries do CDS e Volatilidade Implícita da firma AIG, no tempo e uma versus a outra.

Uma extensão que deve ser considerada em análises é a generalização de (6.22)

$$\Delta u_t = I_t \rho_1 u_{t-1} + (1 - I_t) \rho_2 u_{t-1} + \sum_{i=1}^{p-1} \gamma_i \Delta u_{t-i} + e_t$$

para melhor capturar a dinâmica da correção de Δu_t na direção do seu valor de equilíbrio de longo prazo. Notemos que se $\rho_1 = \rho_2$, recuperaremos a especificação de EG87. Outras extensões existem, ver, por exemplo, Granger, C. W. J. e Teräsvirta, T. (1993).

Para determinar a ordem ótima p , podemos usar o critério AIC. As suposições feitas sobre $\{e_t\}$ devem ser verificadas e testes de diagnóstico podem ser vistos em Granger, C. W. J. e Teräsvirta, T. (1993), Lukkonen, R. et al. (1998), Eitrheim, A. e Teräsvirta, T. (1996), entre outros.

Uma outra alternativa é considerar Δu_{t-1} na especificação da função indicadora da fórmula (6.23), isto é,

$$I_t = \begin{cases} 1, & \text{se } \Delta u_{t-1} \geq \tau \\ 0, & \text{se } \Delta u_{t-1} < \tau \end{cases}, \quad (6.25)$$

denominado modelo M-TAR (*momentum threshold autorregressive*). Também é possível, assim como no modelo TAR, incluir defasagens de Δu_{t-1} .

Enders, W. e Siklos, P. L. (2001) propuseram um teste para detectar cointegração assimétrica. Assim como no teste de EG87, a hipótese nula é de que o vetor \mathbf{x}_t é não cointegrado. Aqui, além da não cointegração, testa-se a não existência do limiar τ e se ρ_1 e/ou ρ_2 é zero. Os testes propostos consideram dois casos. Primeiro $\tau = 0$. O segundo caso considera τ desconhecido.

Caso 1: $\tau = 0$.

O teste baseia-se em estatísticas t e F obtidas na estimação de (6.22) nos modelos TAR e MTAR. As duas estatísticas t testam as hipóteses nulas de $\rho_1 = 0$ e $\rho_2 = 0$. A estatística F testa a hipótese composta $\rho_1 = \rho_2 = 0$. O máximo das duas estatísticas t é chamado de t -max e a estatística F é chamada de Φ . Notemos que ambos ρ_1 e ρ_2 devem ser negativos, assim t -max deve ser negativo. O teste Φ é mais poderoso, mas deve ser usado apenas quando ambas as estimativas de ρ_1 e de ρ_2 são negativas. Enders, W. e Siklos, P. L. (2001) fornecem tabelas de valores críticos dessas estatísticas para alguns tamanhos de amostra. Se o valor estimado para Φ for maior que o valor tabelado, rejeitamos $H_0 : \rho_1 = \rho_2 = 0$. Se o valor de t -max for menor que o valor da tabela, rejeitamos a hipótese nula de não cointegração. Valores críticos para diversos tamanhos de amostras são dados em Enders, W. e Siklos, P. L. (2001). O poder do teste de Enders, W. e Siklos, P. L. (2001) é ligeiramente menor que o do teste de EG87.

Para efetuar os testes t -max e Φ siga os seguintes passos: 1) Efetue a regressão de uma das variáveis em uma constante, e nas outras variáveis (ou outra) (6.19). Obtenha os resíduos $\{\hat{u}_t\}$. Escolha se vai ajustar um modelo TAR ou MTAR (6.23 ou 6.25), mas fixe $\tau = 0$. Estime o modelo de regressão (6.22) e obtenha as estimativas de t -max e Φ . Compare esses valores com os valores críticos dados em Enders, W. e Siklos, P. L. (2001). 2) Se a hipótese alternativa é aceita, pode-se testar se $\rho_1 = \rho_2$. Veja Enders, W. e Falk, B. (1999), onde são usados intervalos de confiança baseado em *bootstrap*. 3) Devemos especificar se a série de resíduos $\{\hat{e}_t\}$ pode ser considerada um RB. Para isto, temos testes e gráficos dados no Capítulo 1. Se, por acaso, os resíduos forem correlacionados, inclua defasagens de Δu_t em (6.22). Isto é, inclua $\gamma_1 \Delta \hat{u}_{t-1}, \gamma_2 \Delta \hat{u}_{t-2}, \dots, \gamma_p \Delta \hat{u}_{t-p}$ considerando ou o modelo TAR ou o MTAR. O número de defasagens pode ser o mesmo utilizado no teste de EG87.

Caso 2: τ desconhecido.

Em geral τ é desconhecido e deve ser estimado. Os valores críticos dos testes desenvolvidos por simulação podem ser vistos em Enders, W. e Siklos, P. L. (2001), a partir de resultados de Chan, K. (1992). Novamente estime (6.19) e obtenha $\{\hat{\mu}_t\}$. Cada resíduo será um possível limiar, mas Enders, W. e Siklos, P. L. (2001) sugerem não considerar os 15% maiores e os 15% menores (ou algum outro percentual). Seria interessante se fazer os gráficos de dispersão dos dados das regressões a serem feitas para conferir este percentual. Veja também outras sugestões de figuras na ilustração da próxima seção. Os restantes \hat{u}_t são considerados possíveis limiares. Para cada limiar, estimamos

(6.22) em um modelo TAR ou MTAR (no caso MTAR ordenamos $\Delta\hat{u}_t$). Aquele limiar \hat{u}_t para o qual a regressão forneceu a menor soma dos quadrados dos resíduos (SQR) será o limiar ótimo. Damos na seção de ilustração uma outra sugestão para escolha do limiar. Novamente valores defasados de $\Delta\hat{u}_t$ podem ser usados se necessários. Chan, K. (1992) mostrou que esta metodologia resulta em uma estimativa super consistente do limiar, mas a distribuição assintótica multivariada dos estimadores dos parâmetros ρ_1, ρ_2 e τ ainda não foi estabelecida.

Existem extensões para esses modelos multivariados que consideram dois, ou três limiares, isto é, vários regimes, veja Balke, N. S. e Fomby, T. (1997).

A partir dos artigos de Granger, C. W. J. (1981), Granger, C. W. J. (1986) e Engle, R. F. e Granger, C. W. J. (1987), ficou clara a importância do conceito de cointegração para a análise econométrica de séries temporais. Como prova disto, temos os inúmeros artigos teóricos e aplicados que se seguiram. Veja por exemplo, artigos publicados em Lenum, M. O. (1992) e Johansen, S. (1995).

6.4 Ilustração

Cointegração tem a ver com simultaneidade de séries temporais, é problema multivariado por natureza. A globalização econômica e a crescente velocidade da internet são bastante responsáveis pela também corrente crescente integração dos mercados. Efeitos são sentidos quase instantaneamente o que torna cada vez mais necessária a consideração conjunta de variáveis, procurando entender e prever seus movimentos simultâneos. Particularmente, em finanças, investidores possuem ativos em diferentes mercados, e é de interesse entender como eles se movem juntos.

A Figura 6.1 mostra à esquerda as séries I(1) do *Credit Default Swap*, CDS, da firma de resseguro AIG e da sua Volatilidade Implícita ao longo do tempo, de 6 de maio de 2002 a 19 de abril de 2012, ao todo 2318 observações. A linha tracejada vertical representa o dia 19 de julho de 2007, dia que podemos considerar que divide os períodos pré- e pós-crise. As séries parecem ser não estacionárias tipo raiz unitária, apesar de não exibirem uma tendência determinística clara. À direita temos uma série versus a outra.

Começaremos ajustando os modelos vistos para cointegração para as três variáveis CDS, Volatilidade Implícita (Vollmp) para a firma AIG, e o índice VIX. A Figura 6.2 mostra as três séries no tempo, para o mesmo período acima.

Os três testes, ADF, PP, e PP modificado (MPP) não rejeitaram a H_0 de RU para Vollmp e VIX ao nível de significância de 5%, porém para o CDS, a hipótese nula foi aceita ao nível de 1% mas rejeitada aos 5%. Utilizamos a metodologia de seleção automática do número de defasagens na utilização da função `unitroot` do SPlus para as estatísticas ADF e MPP conforme Ng, S. e Perron, P. (1995). Continuaremos a análise considerando as três séries I(1) ao nível de 1%.

Aplicando o teste de EG87, primeiro estimamos o vetor β normalizado através da regressão por mínimos quadrados, sendo o CDS a variável dependente, e as outras duas as explanatórias mais uma constante (isto é, na parte determinística da regressão não consideramos uma tendência determinística), obtendo estimativas altamente significativas com R^2 de 76%. Aos resíduos desta regressão aplicamos os testes de RU que neste caso é *sem* termos determinísticos, pois os mesmos já foram considerados na regressão, mas pode ter termos defasados. O número de defasagens no teste RU foi

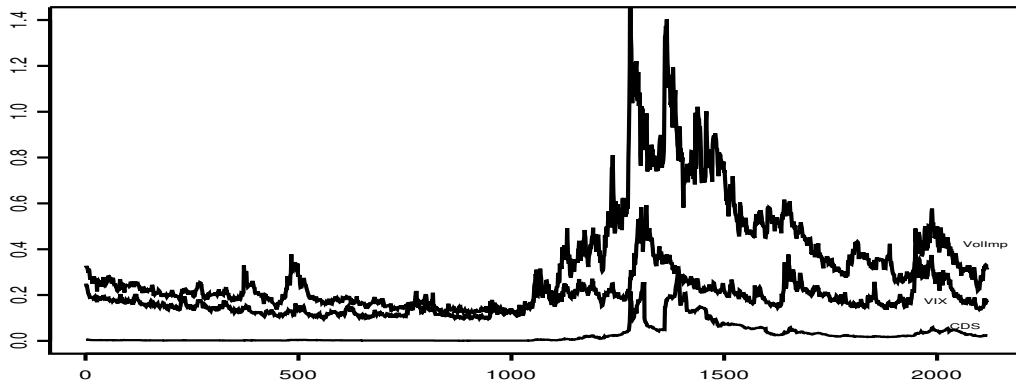


Figura 6.2: Séries do CDS (em baixo), Volatilidade Implícita (alto) e VIX (meio).

determinado conforme a teoria vista para as estatísticas. A suposição feita sobre a parte determinística afeta a distribuição da estatística teste, a distribuição PO, que está tabelada no SPLUS e no R. A H_0 foi fortemente rejeitada para todas as estatísticas de teste e concluímos que existe uma relação de cointegração.

A seguir aplicamos o teste de Johansen para decidir sobre r , o número de vetores de cointegração. No primeiro passo decidimos sobre o número de defasagens no VECM. Utilizando a função `VAR` e o critério AIC identificamos $p = 2$. Em seguida usamos a função `coint` para determinar r . Conforme vimos, é importante especificar corretamente os termos determinísticos na função `coint`. Dentre os cinco casos vistos de tendências e constantes, decidimos pela constante restrita (2), restrita a estar somente no vetor de cointegração, pois as séries não possuem tendência. Abaixo temos o output desta função onde `**` e `*` significam, respectivamente, rejeição a 1% e 5% da hipótese nula ao lado. As duas estatísticas LR_{trace} e LR_{max} aceitam a 1% a $H_0 : r = 2$, isto é, que existem dois vetores de cointegração. Os três auto-valores estimados são 0.0371, 0.0130, e 0.0047, e o número de vetores de cointegração é igual ao número de auto-valores não nulos. Temos então $1 = d - r = 3 - 2$ *tendências estocásticas*.

Hip. Nula	Autovalores	Trac-Est	95%VC	99%VC	Max-Est	95%VC	99%VC
H(0)+++*	0.0371	128.75	34.91	41.07	87.58	22.00	26.81
H(1)+++*	0.0130	41.16	19.96	24.60	30.27	15.67	20.20
H(2)+ *	0.0047	10.89	9.24	12.97	10.89	9.24	12.97

A estimativa da matriz Π de impacto no longo prazo é

	CDS	VolImp	VIX
cds.t	-0.0416	0.0082	-0.0045
VolImp	0.0042	-0.0142	0.0120
VIX	-0.0118	0.0085	-0.0301

As estimativas de máxima verossimilhança do vetor β' normalizado,

$$\begin{aligned}\beta'_1 &= (1, -\beta_{12} - \beta_{13}) \\ \beta'_2 &= (1, -\beta_{22} - \beta_{23})\end{aligned}$$

e da constante restrita ρ_0 são

CDS	VolImp	VIX	Constante restrita
1	-0.2476	0.2410	0.0158
1	0.1092	-0.9418	0.1156

E as relações de cointegração

$$\beta'_t \mathbf{x}_t \sim I(0)$$

estimadas são

$$CDS_t - 0.2476VolImp_t + 0.2410VIX_t + 0.0158 \sim I(0)$$

$$CDS_t + 0.1092VolImp_t - 0.9418VIX_t + 0.1156 \sim I(0)$$

A interpretação é que se as séries de CDS, VolImp, e VIX se afastarem muito das *relações de equilíbrio* $CDS_t = 0.2476VolImp_t - 0.2410VIX_t - 0.0158$ e $CDS_t = -0.1092VolImp_t + 0.9418VIX_t - 0.1156$, forças econômicas irão agir fazendo com que as relações de equilíbrio sejam restauradas.

A matriz α normalizada é

$$\begin{array}{cc} -0.03599 & -0.00488 \\ 0.03651 & -0.01511 \\ -0.02873 & 0.02148 \end{array}$$

Temos um modelo completo, com 3 variáveis ($d = 3$) e 2 relações de cointegração ($r = 2$). O modelo VECM restrito a ser ajustado é

$$\Delta \mathbf{x}_t = \alpha(\beta' \mathbf{x}_{t-1} + \rho_0) + \Gamma_1 \Delta \mathbf{x}_{t-1} + \varepsilon_t$$

Utilizando as estimativas de máxima verossimilhança fornecidas por `coint` podemos escrever

$$\Delta \mathbf{x}_t = \begin{bmatrix} -0.0360 & -0.0049 \\ 0.0365 & -0.0151 \\ -0.0287 & 0.0215 \end{bmatrix} \left(\begin{bmatrix} 1 & -0.2476 & 0.2410 \\ 1 & 0.1092 & -0.9418 \end{bmatrix} \mathbf{x}_{t-1} + \begin{bmatrix} 0.0158 \\ 0.1156 \end{bmatrix} \right) + \Gamma_1 \Delta \mathbf{x}_{t-1} + \varepsilon_t$$

Utilizando agora a função VECM obtemos as estimativas finais de máxima verossimilhança dos coeficientes do modelo completo VECM proposto, isto é, as estimativas das 3 equações

$$\Delta CDS_t = a_1 + \lambda_{11}RC_1 + \lambda_{12}RC_2 + \Gamma_{11}\Delta CDS_{t-1} + \Gamma_{12}\Delta VolImp_{t-1} + \Gamma_{13}\Delta VIX_{t-1} + \varepsilon_{1t}$$

$$\Delta VolImp_t = a_2 + \lambda_{21}RC_1 + \lambda_{22}RC_2 + \Gamma_{21}\Delta CDS_{t-1} + \Gamma_{22}\Delta VolImp_{t-1} + \Gamma_{23}\Delta VIX_{t-1} + \varepsilon_{2t}$$

$$\Delta VIX_t = a_3 + \lambda_{31}RC_1 + \lambda_{32}RC_2 + \Gamma_{31}\Delta CDS_{t-1} + \Gamma_{32}\Delta VolImp_{t-1} + \Gamma_{33}\Delta VIX_{t-1} + \varepsilon_{3t}$$

onde RC_1 e RC_2 representam o primeiro e o segundo resíduo de cointegração.

O *output* da função VECM é

	cds.t	VolImp	VIX
coint.1	-0.0409	0.0214	-0.0072
(std.err)	0.0048	0.0297	0.0127
(t.stat)	-8.4431	0.7199	-0.5702
coint.2	0.0084	-0.0107	0.0095
(std.err)	0.0011	0.0069	0.0029
(t.stat)	7.5033	-1.5600	3.2292
cds.t.lag1	0.1653	-0.1363	-0.1111
(std.err)	0.0202	0.1241	0.0530
(t.stat)	8.1804	-1.0987	-2.0948
VolImp.lag1	0.0171	0.0414	-0.0079
(std.err)	0.0038	0.0232	0.0099
(t.stat)	4.5428	1.7882	-0.7940
VIX.lag1	0.0415	0.1112	-0.1221
(std.err)	0.0084	0.0517	0.0221
(t.stat)	4.9308	2.1525	-5.5285
Intercept	-0.0011	-0.0012	0.0020
(std.err)	0.0003	0.0016	0.0007
(t.stat)	-4.3101	-0.7177	2.9488

Por exemplo, a primeira equação fica:

$$\Delta CDS_t = -0.0011 - 0.0409RC_1 + 0.0084RC_2 + 0.1653\Delta CDS_{t-1} + 0.0171\Delta VolImp_{t-1} + 0.0415\Delta VIX_{t-1}$$

com todas as estimativas dos parâmetros altamente significativas.

A Figura 6.3 mostra os erros de desequilíbrio $\beta' \mathbf{x}_t$ estimados, isto é, as estimativas dos desvios estocásticos das duas relações de equilíbrio. Testamos se as médias dessas séries são zero e aceitamos a H_0 .

Iremos agora ajustar os modelos VECM com limiar para as mesmas variáveis, não só no período completo, mas também separadamente nos períodos pré- e pós crise, períodos esses definidos pela data de 19 de julho de 2007, conforme ilustrado na Figura 6.1.

O teste de EG87 aplicado aos pares (lembramos que os resultados deste teste não são confiáveis quando o modelo certo é o com limiar) e utilizando o período completo indicou cointegração para todos eles. Separadamente, para o período pré-crise e o período pós-crise vimos que

CDS e VolImp são cointegrados no período pré.

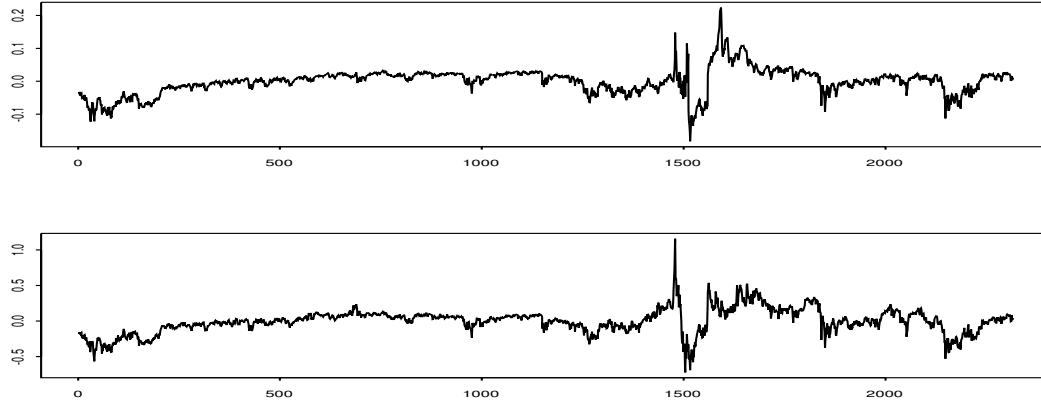


Figura 6.3: Os erros de desequilíbrio para o modelo clássico linear.

- CDS e VolImp são cointegrados no período pós.
- CDS e VIX não são cointegrados no período pré a 1%.
- CDS e VIX não são cointegrados no período pos a 1%.
- VolImp e VIX são cointegrados no período pré a 1%.
- VolImp e VIX não são cointegrados no período pré a 1%.

Novamente estimamos por MQO a regressão, obtendo os resíduos $\{\hat{u}_t\}$ e testamos se são estacionários de raiz unitária. Para o período todo rejeitamos H_0 a qualquer nível de significância razoável. Esses resíduos são usados para estimar o modelo

$$\Delta \hat{u}_t = \rho \hat{u}_{t-1} + \gamma_1 \Delta \hat{u}_{t-1} + \gamma_2 \Delta \hat{u}_{t-2} + e_t$$

que forneceu estatísticas altamente significativas (veja tabela) e AIC de -16615.77. O número de defasagens foi obtidos através do critério AIC. Este é o modelo clássico, linear.

Na segunda coluna da tabela acima fornecemos as estimativas do modelo TAR com limiar $\tau = 0$, equações (6.22) e (6.23) incluindo duas defasagens. A regressão forneceu estimativas altamente significativas. Os valores de ρ_i sugerem convergência e satisfazem as condições de Petrucelli, J. e Woelford, S. (1984). Os valores de Φ e de t-max=-5.35 rejeitam a hipótese nula de não cointegração, respectivamente, $\rho_1 = \rho_2 = 0$ e $\rho_i = 0$. A distribuição da estatística teste depende do tamanho da amostra e do número de defasagens e foi dada na Tabela 3 de Enders, W. e Siklos, P. L. (2001). O AIC foi maior (pior) do que o do modelo linear. Notemos que ρ_1 e ρ_2 estão próximos, mas a estatística F rejeitou a hipótese de serem iguais. A análise dos resíduos deste ajuste não revelou nenhuma irregularidade quanto à suposição de serem um processo RB. No caso de serem esses resíduos correlacionados, deve-se voltar na especificação da regressão acima e incluir outras defasagens de $\Delta \hat{u}_t$.

PERÍODO COMPLETO				
	MODELO CLASSICO	TAR($\tau=0$)	MTAR($\tau=0$)	TAR ($\tau=-0.4$)
ro_1	-0.0527 (-8.28)	-0.0497 (-6.48)	-0.0415 (-4.95)	-0.0430 (-6.44)
ro_2	---	-0.0591 (-5.35)	-0.0675 (-7.02)	-0.1331 (-7.20)
gamma_1	0.0630 (3.04)	0.0632 (3.05)	0.0587 (2.82)	0.0716 (3.46)
gamma_2	0.0733 (3.53)	0.0732 (3.53)	0.0708 (3.41)	0.0798 (3.86)
AIC	-16615.77	-16614.26	-16617.95	-16635.1
Phi	---	19.43	20.38	24.83
t-max	---	-5.35	-4.95	-6.44

A seguir, ajustamos o modelo MTAR com $\tau = 0$. De novo as estimativas dos parâmetros regressão são significativas e as conclusões são as mesmas do ajuste do modelo TAR. O AIC é menor, indicando que este modelo seria melhor.

Agora utilizamos o método de Chan, K. (1992) e estimamos τ a partir dos resíduos ordenados. A necessidade de desconsiderar um percentual $k\%$ dos menores e $k\%$ dos maiores resíduos é na intenção de se viabilizar a regressão, de tal modo que se tenha dados suficientes para um ajuste abaixo e acima do limiar. Entretanto, a decisão sobre $k\%$ é crucial e difícil! Aachamos bastante subjetiva e arbitrária a escolha de $k=15\%$ sugerida em Enders, W. e Siklos, P. L. (2001), e propomos a seguinte análise de sensibilidade.

A Figura 6.4 tem no eixo horizontal os valores possíveis para os limiares, os resíduos ordenados, e no eixo vertical o valor da SQR (SRR) correspondente. As linhas verticais selecionam alguns limiares e marcam os valores correspondentes para o percentual $k\%$. Notamos uma região de estabilidade no centro, quando proporções similares de dados são alocadas para cada regressão. Notamos também muita irregularidade nas duas pontas, podendo-se ter soluções ótimas positivas ou negativas, dependendo de k .

A figura sugere que não deveríamos utilizar o $k=0.5\%$, por apresentar muita instabilidade. Na figura são as as linhas verticais azuis, que corresponderiam a uma solução ótima τ ligeiramente menor que -0.05 . No nosso caso onde o tamanho da amostra é 2318, tomar $k=0.5\%$ significa que a função indicadora vai deixar apenas 11 observações em um dos lados, o que é realmente pouco.

As linhas verticais rosa e verde representam, respectivamente, $k=1\%$ e $k=2\%$. Para 1% a solução é positiva, perto de 0.10 ! A partir daí, a solução será negativa mas cada vez maior. Devemos notar que aqui, 2% são 47 dados, e que 1% separa 24 observações, o que não é pouco. Entretanto a solução ótima é completamente diferente se tomamos $k=1$ ou $k=2$. Com $k=1\%$ o limiar solução é positivo, e com $k=2\%$ o limiar solução é negativo. Para k maiores que 2% o limiar ótimo será sempre negativo, mas diferente. A título de curiosidade colocamos os limites definidos por $k=15\%$ conforme sugerido em Enders, W. e Siklos, P. L. (2001) em linhas tracejadas. A figura sugere ser esta uma solução muito conservativa.

Decidimos que a solução ótima para o limiar seria aquela dada por $k = 2\%$. O limiar é então -0.04033897 e o SQR é 0.1021993 . O resultado seria bem diferente, -0.018 , se tivéssemos considerado apenas os 70% dos resíduos (15% em cada cauda). Considerando este limiar, outros resultados são

	coef	std.err	t.stat	p.value
rho_1	-0.0430	0.0067	-6.4381	0.0000
rho_2	-0.1331	0.0185	-7.1997	0.0000
y (-1)	0.0716	0.0207	3.4613	0.0005
y (-2)	0.0798	0.0207	3.8575	0.0001

com SQR 0.1021993 e AIC = -16643.66 bem menor

Os valores de ρ_i acima sugerem convergência e satisfazem as condições de Petrucelli, J. e Woolford, S. (1984). Esses valores sugerem um ajustamento mais rápido das discrepâncias negativas em relação ao equilíbrio de longo prazo, do que para os positivos. Veja o gráfico dos resíduos (os \hat{u}_t) na figura.

Comparando com a Tabela 1 de Enders, W. e Siklos, P. L. (2001) o valor de Φ é maior que o valor crítico a 1%, e portanto rejeitamos a H_0 de $\rho_1 = \rho_2 = 0$. O valor de t-max (-6.44) comparando na Tabela 2 desses autores ambos rejeitam a hipótese nula de não cointegração. A estatística F padrão rejeita a hipótese de que os coeficientes ρ_1 e ρ_2 são iguais. Os resíduos deste ajuste parecem seguir as suposições feitas sobre eles.

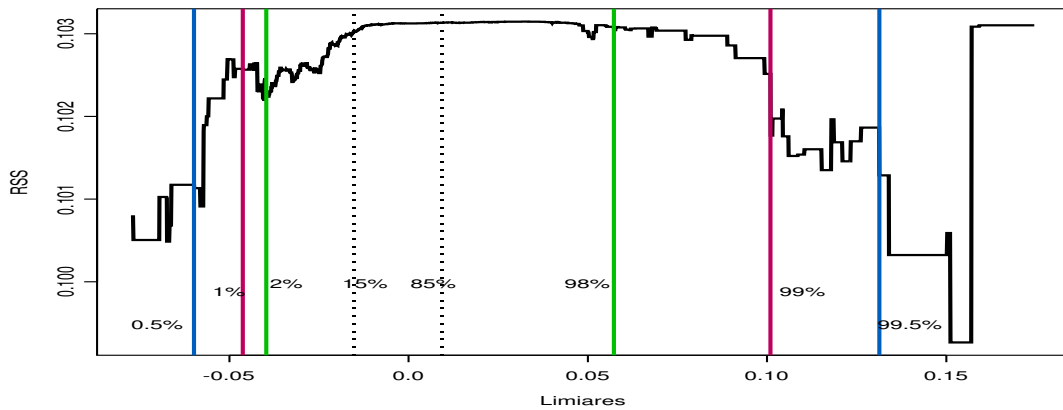


Figura 6.4: *Limiares versus SQR. Linhas verticais representam os percentuais dos resíduos ordenados definindo os possíveis limiares.*

Além desta análise de sensibilidade, podemos também obter um conjunto de valores para o limiar ótimo, todos fornecendo o mesmo valor de SQR. Basta tomar um intervalo aberto entre os resíduos imediatamente menor e maior que a solução. No nosso caso temos, o conjunto de valores z tais que $-0.04042954 \geq z < -0.04018311$.

Seria interessante agora observar a posição temporal dos resíduos considerados como solução ótima

para as diversas escolhas dos percentuais $k\%$. Isto é, primeiramente grafamos o SQR ao longo do tempo, Figura 6.5. Para este gráfico consideramos como limiares apenas aqueles resíduos ordenados maiores (e menores) que os 2% menores e maiores dados. Notemos que o limiar selecionado ocorreu na data de 22 de setembro de 2008 (uma bola no gráfico), e que a maioria dos menores valores para o SQR ocorreram naquela época, estão próximos no tempo.

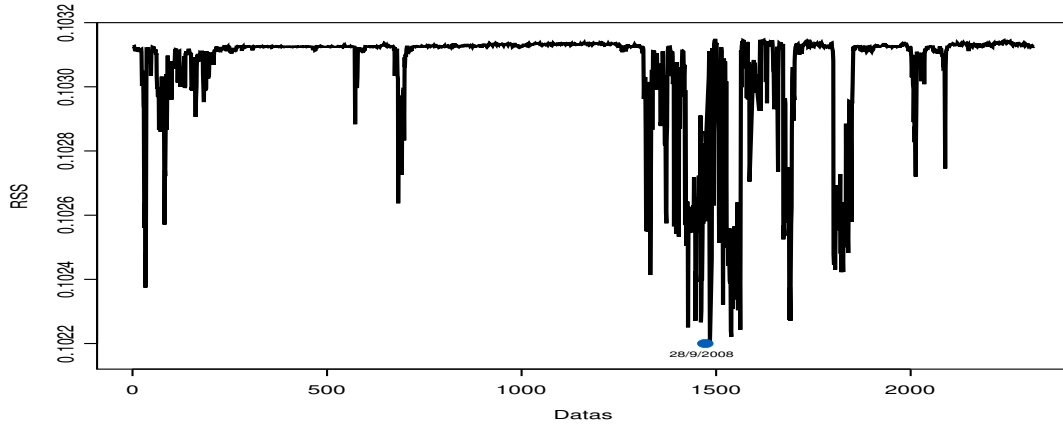


Figura 6.5: *SQR ao longo do tempo. Bola azul indica data do SQR mínimo.*

6.5 Cópulas dinâmicas

A definição dada no Capítulo 3 assumiu que a distribuição d -variada F era constante ao longo do tempo. Por simplicidade seja $d = 2$ e considere o processo estacionário $(X_{1,t}, X_{2,t})_{t \in \mathbb{Z}}$. Quando a lei de probabilidades conjunta de $(X_{1,t}, X_{2,t})$ é independente de t , a estrutura de dependência de (X_1, X_2) é dada pela cópula constante, C , dada por (3.4). Quando ela não é independente de t temos uma cópula com parâmetros variando no tempo, ou, mais do que isto, famílias de cópulas variando no tempo.

Uma cópula com parâmetros variando no tempo pode não satisfazer todas as propriedades de uma cópula. Seja (X_1, X_2) um vetor aleatório contínuo definido em (Ω, \mathcal{A}, P) e assumindo valores em \mathbb{R}^2 , onde \mathcal{A} é a sigma-álgebra gerada por toda a informação passada. No contexto de séries temporais

$$\mathcal{A}_t = \sigma\{x_{1t-1}, x_{2t-1}, (x_{1t-2}, x_{2t-2}), \dots\} \quad t = 1, \dots, T$$

representa a informação passada até o tempo t .

Podemos estender o teorema de Sklar

$$F(x_{1t}, x_{2t} | \mathcal{A}_t) = C_t(F_{1t}(x_{1t} | \mathcal{A}_t), F_{2t}(x_{2t} | \mathcal{A}_t)), \quad (6.26)$$

onde C_t é a cópula no tempo t . Note que em geral a matriz de dados pode não representar T observações da mesma distribuição conjunta. O mesmo conjunto de informação condicional para cada

distribuição univariada e para a cópula condicional garante que a variável transformada é independente da informação contida no conjunto de informação condicional de sua distribuição marginal, garantindo que C_t seja uma cópula.

Patton, A. (2006) introduziu a cópula condicional no caso bivariado ao modelar taxas de câmbio, e assumiu que a transformação logística do parâmetro da cópula Gaussiana seguia um processo ARMA. Rockinger, M. e Jondeau, E. (2001) assumiram uma cópula paramétrica condicional à posição de observações passadas no quadrado unitário, isto combinado com a estimação marginal prévia através de modelos GARCH. Fermanian e Wegkamp (2004) introduziram o conceito de *pseudo-cópulas* e mostraram que os dois trabalhos anteriores tratavam de pseudo-cópulas. Eles propuseram um estimador não-paramétrico para as pseudo-cópulas condicionais, derivaram sua distribuição assintótica e elaboraram um teste de ajuste. Cherubini et al. (2004) ajustaram um modelo GARCH(1,1) à dados financeiros e modelaram a estrutura de dependência com uma cópula Gaussiana onde o coeficiente de correlação variaria no tempo de acordo com um modelo ARMAX. Dias e Embrechts (2004) utilizaram também modelos GARCH e a cópula t ao modelar dados financeiros de alta frequência. Mendes, B. V. M. (2005b) estendeu o modelo de Rockinger, M. e Jondeau, E. (2001) assumindo uma pseudo-cópula condicional à posição de observações passadas de *lag 1*, de *lag 2*, e de suas interações, no quadrado unitário, combinando com a estimação prévia das marginais usando modelos GARCH com memória longa.

Van den Goorbergh, R. W. J., Genest, C. e Werker, B. J. M. (2005) estudaram o comportamento de preços de opções supondo que o ativo principal seguiria uma cópula dinâmica. Eles estimaram o parâmetro da cópula através de uma janela deslizante, e ajustaram um modelo linear simples, onde a variável explicativa seria uma função da volatilidade passada, e onde as observações da variável dependente seriam as estimativas do coeficiente de correlação. Veja também este estudo envolvendo memória longa na volatilidade e o seu efeito na verdadeira cópula dos ativos em Mendes, B. V. M. e Koley, N. (2008).

Um modelo muito interessante para cópulas dinâmicas, desenvolvido por Mendes, B. V. M. e Lopes, S. R. C. (2011), assume que o parâmetro θ da cópula segue um processo ARFIMA(p, d, q). Neste modelo foi desenvolvido o conceito de cópulas com memória longa. Neste artigo demonstramos a existência de memória longa na estrutura de dependência, ou na cópula. Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo ARFIMA(p, d, q). Neste caso

$$\Phi(\mathcal{L})(1 - \mathcal{L})^d X_t = \Theta(\mathcal{L})\epsilon_t, \quad d \in \mathbb{R}, \quad (6.27)$$

onde \mathcal{L} é o operador retroativo, isto é, $\mathcal{L}^k X_t = X_{t-k}$. Os polinômios $\Phi(\mathcal{L}) = \sum_{i=0}^p (-\phi_i) \mathcal{L}^i$ e $\Theta(\mathcal{L}) = \sum_{j=0}^q (-\theta_j) \mathcal{L}^j$ possuem graus p e q , respectivamente, com $\phi_0 = -1 = \theta_0$. O processo $\{\epsilon_t\}_{t \in \mathbb{Z}}$ é um ruído branco com média zero e variância finita σ_ϵ^2 . O termo $(1 - \mathcal{L})^d$ é a expansão binomial em \mathcal{L} .

O processo $\{X_t\}_{t \in \mathbb{Z}}$ dado pela expressão (6.27) é chamado de *processo geral fracionalmente diferenciado com média zero*, e o parâmetro d é o *parâmetro de diferenciação fracional*. Assim definido, este processo será estacionário e inversível se as raízes de $\Phi(\cdot)$ e $\Theta(\cdot)$ estiverem fora do círculo unitário e $|d| < 0.5$. A função densidade espectral, $f_X(\cdot)$, é dada por

$$f_X(w) = f_U(w) \left(2 \sin\left(\frac{w}{2}\right) \right)^{-2d}, \quad w \in [-\pi, \pi], \quad (6.28)$$

onde $f_U(\cdot)$ é a função densidade espectral de um processo ARMA(p, q). Notemos que $f_X(w) \simeq w^{-2d}$, quando $w \rightarrow 0$. O processo ARFIMA(p, d, q) possui *memória longa* quando $d \in (0.0, 0.5)$, e possui memória intermediária quando $d \in (-0.5, 0.0)$ e *memória curta* quando $d = 0$.

Seja $\delta \in \Delta \subset \mathbb{R}$ o parâmetro da cópula. O subconjunto Δ irá definir a transformação apropriada $g(X_t)$, $g : \mathbb{R} \mapsto \Delta$ de tal forma que $g(X_t) = \delta_t \in \Delta$.

Um detalhe interessante é que é possível que este tipo de dependência exista *apenas* na cópula, e que as distribuições marginais possuam apenas memória curta ou mesmo nenhuma estrutura temporal.

Quando a cópula pertence à uma família paramétrica, $\{C_\theta, \theta \in \Theta\}$, estimativas para os parâmetros podem ser obtidas em dois passos. Tanto o primeiro passo (estimação marginal) e o segundo (estimação da cópula) podem ser feitos parametricamente (máxima verossimilhança, mínimos quadrados, mínima distância, estimação robusta) ou semi-parametricamente. Veja detalhes em Genest, C. et al. (1993), Shih, J. H. e Louis, T. A. (1995), Joe, H. (1997), Mendes, B. V. M. et. al. (2007), Vandenhende e Lambert (2005), etc. O método de estimação por máxima verossimilhança *local* foi adaptado para fazer inferências baseadas em cópulas dinâmicas em Mendes, B. V. M. e Melo, E. F. (2009) e Mendes, B. V. M. e Melo, E. F. (2010).

Em finanças, o primeiro passo (a estimação das distribuições marginais) pode ser feita com bastante sucesso pela grande classe dos modelos ARFIMA-FIEGARCH. Os resíduos padronizados são então utilizados para estimar dinamicamente a estrutura de dependência. Muitas vezes ainda existe alguma estrutura dinâmica residual após os ajustes marginais justificando a estimação dinâmica da cópula pertinente.

Finalmente, vale a pena notar que esta modelagem permite estudar as variações no tempo de toda a estrutura de dependência, e não apenas de um coeficiente de correlação.

6.6 Exercícios

Exercício 1: Escreva a função de log verossimilhança para a estimação dos parâmetros do VAR irrestrito em notação matricial. Mostre que $\ln L_{\max} = -\frac{1}{2}T \ln |\hat{\Sigma}| + \text{constantes}$.

Exercício 2: Ache a expressão para τ_0 na representação de Granger-Johansen quando $k = 1$.

Exercício 3: Gere duas variáveis I(1) cointegradas segundo um modelo TAR e faça o gráfico delas no tempo. Sugestão: Suponha que o vetor de cointegração seja tal que o sistema esteja no seu equilíbrio de longo prazo sempre que $x_{1,t} = x_{2,t}$. Para representar as sequências $\{\nu_{1,t}\}$ e $\{\nu_{2,t}\}$ gere dois conjuntos de 500 v.a.'s i.i.d. $\sim N(0,1)$. Sejam $\rho_{1,1} = -\rho_{1,2} = -0.05$ e $\rho_{2,1} = -\rho_{2,2} = -0.25, \tau = 0$, e valores iniciais iguais a zero. Então use (6.24) para gerar 500 $\{x_{1,t}\}$ e $\{x_{2,t}\}$. Observe as velocidades de ajustamento.

Exercício 4: No exercício anterior, calcule o percentual(%) de discrepâncias observadas positivas persistentes de um período para outro, e o % das negativas. O % positivo é maior?

Exercício 5: Troque os valores de $\rho_{1,\cdot}$ e $\rho_{2,\cdot}$ do exercício 3 e repita os exercícios 3 e 4.

6.7 Apêndice do Capítulo 6

AP6.1: O Modelo VECM(2).

Considere o modelo VAR(2):

$$\mathbf{x}_t = \Phi \mathbf{D}_t + \Pi_1 \mathbf{x}_{t-1} + \Pi_2 \mathbf{x}_{t-2} + \varepsilon_t$$

ou

$$(\mathbf{I} - \Pi_1 B - \Pi_2 B^2) \mathbf{x}_t = \Phi \mathbf{D}_t + \varepsilon_t$$

Logo, as raízes de $|\Pi(z)| = |\mathbf{I} - \Pi_1 z - \Pi_2 z^2|$ fornecem informação sobre a estacionaridade de \mathbf{x}_t . Se as raízes de $|\Pi(z)|$ estão fora do círculo unitário, então \mathbf{x}_t é estacionário. Se algumas raízes estão fora e algumas no círculo unitário, então \mathbf{x}_t é não estacionário. E se qualquer uma das raízes está dentro do círculo unitário, então \mathbf{x}_t é explosivo. Note que podemos ainda achar raízes resolvendo para os autovalores da matriz associada, que são iguais a z^{-1} .

Se Π tiver posto reduzido $r \leq d$, pode ser escrito como

$$\Pi = \alpha \beta'$$

onde α e β são matrizes $d \times r$ de posto cheio. Então:

$$\Delta \mathbf{x}_t = \Phi \mathbf{D}_t + \alpha \beta' \mathbf{x}_{t-1} + \Gamma_1 \Delta \mathbf{x}_{t-1} + \dots + \Gamma_{p-1} \Delta \mathbf{x}_{t-p+1} + \varepsilon_t \quad (6.29)$$

onde α representa a velocidade do ajustamento na direção do equilíbrio e $\beta' \mathbf{x}_{t-1}$ é um vetor $rx1$ de relações de cointegração estacionárias. Assim, todas as variáveis em (6.29) são estacionárias.

Em resumo, existe cointegração se (A) posto(Π) = $r \leq d$. (B) o número de raízes unitárias no sistema é $d - r$; e (C) teremos r relações estacionárias. Para observar (B) e (C), considere um VAR(1) sem componentes determinísticas, e decompõe nos espaços $d - r$ e r :

$$A \mathbf{x}_t = \{A(I_d + \alpha \beta') A^{-1}\} A \mathbf{x}_{t-1} + \varepsilon_t$$

$$\begin{pmatrix} \beta'_\perp \mathbf{x}_t \\ \beta' \mathbf{x}_t \end{pmatrix} = \begin{pmatrix} I_{d-r} & \beta'_\perp \alpha \\ 0 & I_r + \beta' \alpha \end{pmatrix} \begin{pmatrix} \beta'_\perp \mathbf{x}_{t-1} \\ \beta' \mathbf{x}_{t-1} \end{pmatrix} + \begin{pmatrix} \beta'_\perp \varepsilon_t \\ \beta' \varepsilon_t \end{pmatrix},$$

onde $\beta_\perp \in \mathbb{R}^{d \times (d-r)}$ é o complemento ortogonal de β tal que $\beta' \beta_\perp = 0$.

Encontremos as raízes resolvendo o problema de autovalores usando $|\rho I - M| = 0$, onde ρ é um autovalor da matrix quadrada M :

$$\begin{aligned} 0 &= \left| \rho I_d - \begin{pmatrix} I_{d-r} & \beta'_\perp \alpha \\ 0 & I_r + \beta' \alpha \end{pmatrix} \right| \\ &= \left| \begin{pmatrix} (\rho - 1)I_{d-r} & \beta'_\perp \alpha \\ 0 & \rho I_r - (I_r + \beta' \alpha) \end{pmatrix} \right| \\ &= (\rho - 1)^{d-r} | \rho I_r - (I_r + \beta' \alpha) |. \end{aligned}$$

Logo, existem ao menos $d - r$ raízes unitárias. Sobre as raízes de $|\rho I_r - (I_r + \beta' \alpha)|$, dada a suposição (B), existem $d - r$ raízes unitárias no total, e então 1 não pode ser raiz. Isso implica que $|\beta' \alpha| \neq 0$. Dada a suposição (C), existem r raízes estacionárias. Logo, os autovalores absolutos de $(I_r + \beta' \alpha)$ são < 1 .

Para obter a representação de Granger-Johansen, considere as seguintes relações de cointegração: Pré multiplique (6.29) sem constantes por β' :

$$\Delta(\beta' \mathbf{x}_t) = \beta' \alpha \beta' \mathbf{x}_{t-1} + \beta' \varepsilon_t, \quad \text{logo}$$

$$\beta' \mathbf{x}_t = (I_r + \beta' \alpha) \beta' \mathbf{x}_{t-1} + \beta' \varepsilon_t = \sum_{s=0}^{t-1} (I_r + \beta' \alpha)^s (\beta' \varepsilon_{t-s}) + (I_r + \beta' \alpha)' \beta' \mathbf{x}_0$$

é aproximadamente estacionário se os autovalores absolutos de $I_r + \beta' \alpha < 1$.

Considere agora as tendências comuns. Multiplique antes por α'_\perp :

$$\alpha'_\perp \Delta \mathbf{x}_t = (\alpha'_\perp \alpha)' \beta' \mathbf{x}_{t-1} + \alpha'_\perp \varepsilon_t = \alpha'_\perp \varepsilon_t$$

e some para então

$$\alpha'_\perp \mathbf{x}_t = \alpha'_\perp \sum_{s=1}^t \varepsilon_s + \alpha'_\perp \mathbf{x}_0$$

Usando a representação Granger-Johansen de um VAR(1):

$$\mathbf{x}_t = (\alpha(\beta' \alpha)^{-1} \beta' + \beta_\perp (\alpha'_\perp \beta_\perp)^{-1} \alpha'_\perp) \mathbf{x}_t \approx \alpha(\beta' \alpha)^{-1} \quad (\text{processo estacionário}) + \beta_\perp (\alpha'_\perp \beta_\perp)^{-1} (\alpha'_\perp \sum_{s=1}^t \varepsilon_s + \alpha'_\perp \mathbf{x}_0).$$

Dadas as suposições (A), (B) e (C), a representação Granger-Johansen (ou MA) de um VAR(p) é dada por

$$\mathbf{x}_t \approx \mathbf{C} \sum_{s=1}^t \varepsilon_s + \mathbf{C}^*(B) \varepsilon_t + \tilde{\mathbf{X}}_0,$$

onde $\mathbf{C} = \beta_\perp (\alpha'_\perp \Gamma \beta_\perp)^{-1} \alpha'_\perp$ com $\Gamma = -(I - \Gamma_1 - \dots - \Gamma_{p-1})$, $\beta' \tilde{\mathbf{X}}_0 = 0$ e $\mathbf{C}^*(B) \varepsilon_t$ é um processo estacionário.

O modelo VAR(p) pode ser estendido. Considere o VAR(1) com um coeficiente de tendência μ_1 e constante μ_0 :

$$\Delta \mathbf{x}_t = \alpha \beta' \mathbf{x}_{t-1} + \mu_0 + \mu_1 t + \varepsilon_t.$$

Temos que

$$\mu_0 = (\alpha(\beta' \alpha)^{-1} \beta' + \beta_\perp (\alpha'_\perp \beta_\perp)^{-1} \alpha'_\perp) \mu_0 \equiv \alpha \beta_0 + \gamma_0$$

$$\mu_1 = (\alpha(\beta' \alpha)^{-1} \beta' + \beta_\perp (\alpha'_\perp \beta_\perp)^{-1} \alpha'_\perp) \mu_1 \equiv \alpha \beta_1 + \gamma_1$$

Que nos dá

$$\begin{aligned} \Delta \mathbf{x}_t &= \alpha \beta' \mathbf{x}_{t-1} + \alpha \beta_0 + \alpha \beta_1 t + \gamma_0 + \gamma_1 t + \varepsilon_t \\ &= \alpha(\beta', \beta_0, \beta_1) \begin{pmatrix} \mathbf{x}_{t-1} \\ 1 \\ t \end{pmatrix} + \gamma_0 + \gamma_1 t + \varepsilon_t \end{aligned}$$

Existem cinco casos particulares.

1. $\mu_1 = \mu_0 = 0$. Série sem componentes determinísticos.
2. $\mu_1 = \gamma_0 = 0$ mas $\beta_0 \neq 0$. Uma constante restrita apenas nas relações de cointegração.
3. $\mu_1 = 0$ mas μ_0 é irrestrita. Uma constante nas relações de cointegração, e tendência linear nos níveis (\mathbf{x}_t).

4. $\gamma_1 = 0$ mas $(\gamma_0, \beta_0, \beta_1) \neq 0$. Uma tendência restrita somente nas relações de cointegração, e uma constante irrestrita.
5. Sem restrições em μ_0 ou μ_1 . Tendência e constante irrestritas. A tendência será quadrática nos níveis \mathbf{x}_t .

Para obter a representação de Granger-Johansen, invertamos o VAR(1) para chegar na forma MA:

$$\begin{aligned}\mathbf{x}_t &= \mathbf{C} \sum_{i=1}^{\infty} (\varepsilon_i + \mu_0 + \mu_1 i) + \mathbf{C}^*(B)(\varepsilon_t + \mu_0 + \mu_1 t) \\ &= \mathbf{C} \sum_{i=1}^t \varepsilon_i + \mathbf{C} \mu_0 t + \frac{1}{2} \mathbf{C} \mu_1 t + \frac{1}{2} \mathbf{C} \mu_1 t^2 \\ &\quad + \mathbf{C}^*(B) \varepsilon_t + \mathbf{C}^*(B) \mu_0 + \mathbf{C}^*(B) \mu_1 t + \tilde{\mathbf{X}}_0\end{aligned}$$

quando somamos sobre uma amostra finita de 1 até T. Mas

$$\begin{aligned}\alpha'_\perp \mu_0 t &= \alpha'_\perp \alpha \beta_0 t + \alpha'_\perp \gamma_0 t = \alpha'_\perp \gamma_0 t \\ \alpha'_\perp \frac{1}{2} \mu_1 t &= \frac{1}{2} (\alpha'_\perp \alpha \beta_1 t + \alpha'_\perp \gamma_1 t) = \frac{1}{2} \alpha'_\perp \gamma_1 t \\ \alpha'_\perp \frac{1}{2} \mu_1 t^2 &= \frac{1}{2} (\alpha'_\perp \alpha \beta_1 t^2 + \alpha'_\perp \gamma_1 t^2) = \frac{1}{2} \alpha'_\perp \gamma_1 t^2\end{aligned}$$

Então,

$$\mathbf{x}_t = \mathbf{C} \sum_{i=1}^t \varepsilon_i + \mathbf{C} \gamma_0 t + \frac{1}{2} \mathbf{C} \gamma_1 t + \frac{1}{2} \mathbf{C} \gamma_1 t^2 + \mathbf{C}^*(B) \varepsilon_t + \mathbf{C}^*(B) \mu_0 + \mathbf{C}^*(B) \mu_1 t + \tilde{\mathbf{X}}_0$$

Assim, tendências lineares podem originar-se de três formas diferentes no modelo VAR:

1. Do termo $\mathbf{C}^*(B) \mu_1 t$ da tendência linear restrita ou não restrita, $\mu_1 t$.
2. Do termo $\gamma_1 t$ da tendência linear irrestrita de $\mu_1 t$.
3. Do termo $\gamma_0 t$ da constante irrestrita μ_0 .

Compactamente, a representação de Granger-Johansen é dada por:

$$\mathbf{x}_t = \tau_0 + \tau_1 t + \tau_2 t^2 + \mathbf{C} \sum_{i=1}^t \varepsilon_i + \mathbf{C}^*(B) \varepsilon_t + \tilde{\mathbf{X}}_0.$$

Para determinar a ordem p estime o modelo de ordem $p+1$

$$\mathbf{x}_t = A_1 \mathbf{x}_{t-1} + \dots + A_p \mathbf{x}_{t-p} + A_{p+1} \mathbf{x}_{t-p-1} + \varepsilon_t$$

Calcule a razão das verossimilhanças (LR) para testar $A_{p+1} = 0$:

$$LR(\mathcal{H}_p | \mathcal{H}_{p+1}) = -2 \ln \mathcal{Q}(\mathcal{H}_p | \mathcal{H}_{p+1}) = T(\ln |\hat{\Sigma}_p| - \ln |\hat{\Sigma}_{p+1}|), \quad (6.30)$$

onde \mathcal{H}_p é hipótese nula de p lags, enquanto \mathcal{H}_{p+1} é hipótese alternativa que $p+1$ defasagens são necessárias. Logo, $LR \xrightarrow{D} \chi^2(p^2)$.

Teste para autocorrelação dos resíduos

Estime uma regressão cujas as variáveis são os resíduos do VAR estimado com p variáveis defasadas e o resíduo do VAR defasado de j ordens:

$$\hat{\varepsilon}_t = A_1 \mathbf{x}_{t-1} + \dots + A_p \mathbf{x}_{t-p} + A_\varepsilon \hat{\varepsilon}_{t-j} + \tilde{\varepsilon}_t.$$

Queremos que $\hat{\varepsilon}_t \approx \tilde{\varepsilon}_t$, e usamos um teste Multiplicador de Lagrange (LM) (calculado como um teste de razão de Wilk) com uma correção para amostras pequenas:

$$LM(j) = -(T - p(p+1) - \frac{1}{2}) \ln \left(\frac{|\Sigma(j)|}{|\hat{\Sigma}|} \right),$$

o qual é aproximadamente distribuído como uma χ^2 com p^2 graus de liberdade.

Teste para detectar efeito ARCH

Calcule o R^2 da regressão auxiliar

$$\hat{\varepsilon}_{i,t}^2 = \gamma_0 + \sum_{j=1}^m \gamma_j \hat{\varepsilon}_{i,t-j}^2 + u_{i,t}.$$

Se $R^2 = 1 - \frac{\sum u_{i,t}^2}{\sum [\varepsilon_{i,t}^2 - \text{avg}_t(\varepsilon_{i,t}^2)]^2}$ é pequeno, as variâncias são, provavelmente, não correlacionadas.

O teste ARCH de ordem m é calculado como $(T + k - m) \times R^2$, onde T é o tamanho da amostra e p o número de defasagens do VAR, e $(T + p - m) \times R^2 \xrightarrow{D} \chi^2(m)$. Notemos que Rahbek et. al. (2002) mostrou que os testes de cointegração baseados nos postos são robustos contra efeitos moderados tipo ARCH nos resíduos.

Teste para Normalidade

Estime o modelo de ordem p , e verifique os terceiros e quartos momentos dos resíduos (assimetria zero e curtose 3 se normais). Para a assimetria e curtose calcule, respectivamente

$$A_i = T^{-1} \sum_{t=1}^T \left(\frac{\hat{\varepsilon}_i}{\hat{\sigma}_i} \right)_t^3, \text{ e } C_i = T^{-1} \sum_{t=1}^T \left(\frac{\hat{\varepsilon}_i}{\hat{\sigma}_i} \right)_t^4.$$

Conforme vimos no Capítulo 1, a estatística teste é assintoticamente distribuída como

$$\frac{T}{6}(A_i)^2 + \frac{T}{24}(C_i - 3)^2 \stackrel{a}{\sim} \chi^2(2).$$

Capítulo 7

Medidas de Risco

Todos os dias as instituições financeiras estimam medidas de risco. Os números obtidos são utilizados em tomadas de decisão, as quais podem envolver consideráveis somas de dinheiro. Essas estimativas também devem ficar disponíveis para as agências reguladoras, auditores internos e externos, e são o ponto de partida para o cálculo das reservas financeiras para cobrir possíveis perdas. Se a estimativa for muito conservativa (superestimar o risco), pode haver perda no lucro já que um volume desnecessário será colocado à parte. Se, por outro lado, as medidas de risco forem subestimadas, elas serão ultrapassadas com maior frequência e perderão sua função. Tudo isto motiva o aperfeiçoamento do cálculo de medidas de risco, com a utilização de modelos econométricos e modelos para valores extremos.

Neste capítulo vamos calcular algumas medidas de risco. Apesar de ser impossível tratar de todas as propostas existentes na literatura, iremos fornecer uma lista bem completa e atualizada sobre o assunto. Para uma boa discussão sobre risco veja McNeil, A. J., Frey, R. e Embrechts, P. (2005).

Para vetores com distribuição elíptica, medidas de risco clássicas tais com o VaR ou a abordagem de média-variância de Markowitz para alocação de ativos, funcionam bem. Entretanto, fora do mundo elíptico, elas se deterioram: O VaR é tipicamente não sub-aditivo, as alocações obtidas não são consistentes e dependem da medida de risco utilizada, e métodos baseados no coeficiente de correlação linear são insuficientes.

Algumas crenças que não se justificam (falácias) ainda permeiam o mundo financeiro: (1) que basta conhecer as marginais e ρ para se determinar a distribuição multivariada; (2) que dados dois ativos X e Y, o $\text{VaR}(X+Y)$ é máximo quando a correlação entre eles é máxima; (3) que uma correlação bem pequena entre X e Y significa que esses ativos sejam praticamente independentes. A solução para (1) é usar cópulas. Para (2) existem várias publicações com resultados sobre as propriedades do VaR; e (3) é dar um contra-exemplo: use uma cópula com coeficiente de dependência de cauda positivo (por exemplo, uma cópula t com $\rho = 0.00$ e $\nu = 4$).

7.1 Valor em Risco

Considere uma carteira de investimentos formada por d ativos A_1, A_2, \dots, A_d , com pesos w_1, w_2, \dots, w_d respectivamente. Seja r_t o retorno logarítmico percentual associado à esta carteira. Medir o risco desta carteira essencialmente se reduz a estimar a sua distribuição de probabilidade F , $F(x) = P(r_t \leq x)$, ou funções de F como sua variância ou seus quantis de probabilidade acumulada α . Até pouco tempo atrás, os modelos para o cálculo de medidas de risco assumiam que os retornos eram normalmente distribuídos, devido à fácil implementação desta premissa. Essa premissa está, entretanto, distante da realidade e inúmeros estudos empíricos já demonstraram que a distribuição de perdas e ganhos de ações ou carteiras possuem caudas mais pesadas que as previstas pela distribuição Normal.

O Valor-em-Risco (VaR) é provavelmente a medida de risco mais utilizada desde 1996, quando uma emenda ao Acordo da Basiléia I determinou que as instituições financeiras poderiam decidir internamente sobre a metodologia de cálculo do VaR. Essas metodologias seria auditadas pelas agências reguladoras americanas. No Brasil, o Banco Central adotou a mesma medida em 2001 para alguns mercados.

Informalmente, o VaR_α pode ser definido como um valor limite de perda tal que a probabilidade de que uma perda da carteira em um dado horizonte de tempo exceda este valor é α .

Considere novamente a série de retornos r_t , $t = 1, 2, \dots, T$ (pode ser de uma carteira ou de um único ativo) e sua distribuição de probabilidade $F(x)$. Aqui não estamos fazendo distinção se os retornos são de um-período ou de k -períodos ou se a distribuição é condicional ou não condicional. Se quisermos definir uma estatística baseada em F e que meça a severidade do risco de se manter a posição desta carteira durante um horizonte Δ , o candidato mais plausível seria a *perda máxima*, isto é, $\inf\{x \in \Re : F(x) = 0\}$. Contudo, as distribuições de probabilidade utilizadas têm em geral suporte não limitado, e neste caso a perda máxima seria infinita. Além disto, esta estimativa não depende da distribuição F . Mais interessante é substituir a perda máxima por “perda máxima a qual não é excedida com probabilidade $(1 - \alpha)$ ”.

Assim, mais formalmente, o Valor-em-Risco é o menor número x tal que a probabilidade que o retorno exceda x não é maior que α :

$$\text{VaR}_\alpha = \inf\{x \in \Re : P(r_t > x) \leq 1 - \alpha\} = \inf\{x \in \Re : F(x) \geq \alpha\}$$

onde a probabilidade $P(\cdot)$ pode ser condicional ou não condicional. Atenção, *podemos definir o risco na cauda direita*:

$$\text{VaR}_\alpha = \sup\{x \in \Re : F(x) \leq 1 - \alpha\}.$$

Valores típicos para $(1 - \alpha)$ são 0.95, 0.99, 0.999, e o horizonte de tempo Δ para risco de mercado é usualmente 1 período, ou, se retornos diários, 1 ou 10 dias. Estatisticamente, o VaR nada mais é que o quantil de uma distribuição. Lembremos a definição da função quantil, que fornece o q -quantil,

r_q , da distribuição F de r_t :

$$r_q = \inf\{x \in \mathfrak{R} : F(x) \geq q\} = F^{-1}(q) .$$

Assim, para o cálculo do VaR_α de um ativo precisamos calcular $F^{-1}(q)$, sendo $q = \alpha$ se na cauda esquerda, ou $q = (1 - \alpha)$ se for a cauda direita da distribuição. Iremos por simplicidade, e também devido aos modelos da TVE, sempre falar da cauda direita, e neste caso α é a *probabilidade de excedência*. Na prática, ao se estimar o VaR de uma posição comprada, multiplica-se os dados por (-1) e se considera a cauda direita.

O VaR tem sido criticado por não ser sub-aditivo (veja Embrechts, P., McNeil, A. J. e Straumann, D. (2002) e McNeil, A. J., Frey, R. e Embrechts, P. (2005) para um panorama geral deste assunto). Isto significa que o risco de uma carteira, quando este é medido pelo VaR, pode ser maior que a soma dos riscos individuais de cada componente (Artzner, P., Delbaen, F., Eber, J. e Heath, D. (1997) e Artzner, P., Heath, D., Delbaen, F. e Eber, J.M. (1998)). Assim, gerenciar o risco de uma carteira através do VaR pode não estimular a diversificação. Mesmo assim, o acordo final do *Basle Capital Supervision (2005)*.

Diferentes formas de estimar a F nos levam a valores diferentes do VaR. Iremos agora estimar o VaR condicional e não condicional utilizando várias metodologias e comparar suas *performances* através do número de violações e testes.

7.2 VaR Não Condicional

Os diversos métodos para o VaR Não Condicional são:

1. VaR Não Condicional Empírico
2. VaR Não Condicional Baseado no Modelo Normal
3. VaR Não Condicional Baseado no Modelo Normal Assimétrico
4. VaR Não Condicional Baseado no Modelo t-student
5. VaR Não Condicional Baseado no Modelo t-student Assimétrico
6. VaR Não Condicional Baseado no Modelo GEV
7. VaR Não Condicional Baseado no Modelo GEV e Índice Extremal θ
8. VaR Não Condicional Baseado no Modelo r -maiores
9. VaR Não Condicional Baseado no Modelo r -maiores e Índice Extremal θ
10. VaR Não Condicional Baseado no Modelo GPD
11. VaR Não Condicional Baseado no Modelo GPD e Índice Extremal θ

12. VaR Não Condicional Utilizando Cópulas

13. VaR Não Condicional Baseado em Simulações

O VaR não-condicional é um quantil da distribuição não-condicional dos retornos. Seja $r_t \sim F$ o log-retorno diário. De acordo com as diversas escolhas da distribuição F , temos as possibilidades já enumeradas. A Tabela 7.1 ilustra para alguns dos procedimentos e resume os valores obtidos para o VaR de probabilidades 0.001, 0.01, e 0.05 e para as perdas do índice do Brasil.

(1) VaR Empírico.

É a abordagem não-paramétrica para o cálculo do VaR. O VaR_α empírico é o $(1 - \alpha)$ -quantil da distribuição empírica dos dados, F_T :

$$F_T(x) = \frac{1}{T} \sum_{t=1}^T \mathbf{1}_{(r_t \leq x)}$$

para $x \in \mathbb{R}$, e, portanto,

$$\text{VaR}_\alpha = \inf_{r_t \text{ na série}} \{r_t : F_T(r_t) \geq (1 - \alpha)\} = F_T^{-1}(1 - \alpha).$$

Nesse caso, supomos que a distribuição preditiva dos retornos é igual à da amostra. Sob essa abordagem, o VaR de k períodos é igual ao VaR de 1 período. Esse procedimento fornece estimativas ruins para α pequeno (menor que 0.05), e T pequeno.

Tabela 7.1: *Estimativas do VaR Não Condicional de probabilidades 0.001, 0.01, e 0.05 e para as perdas do índice brasileiro.*

Metodologia	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.001$
1.VaR Não Condicional Empírico	-2.8982	-4.3966	-6.6047
2.VaR Não Condicional Baseado no Modelo Normal	-2.7367	-3.9111	-5.2276
4.VaR Não Condicional Baseado no Modelo t-student (25df)	-2.7256	-4.0099	-5.6051
6.VaR Não Condicional Baseado no Modelo GEV	-2.4788	-4.2180	-6.5590
7.VaR Não Condicional Baseado no Modelo GEV e Índice Extremal θ	-2.8667	-4.5900	-6.9096
8.VaR Não Condicional Baseado no Modelo r -maiores	-2.7220	-4.4590	-6.7939
9.VaR Não Condicional Baseado no Modelo r -maiores e Índ. Extremal θ	-3.1096	-4.8304	-7.1434
10.VaR Não Condicional Baseado no Modelo GPD	-2.9016	-4.5235	-6.3230
11.VaR Não Condicional Baseado no Modelo GPD e Índice Extremal θ	-3.0158	-4.6178	-6.3930

Os modelos 6,7,8,9 têm $n = 22$, os modelos 7,9,11 têm θ baseado em $n = 10$, e nos modelos 8 e 9 $r = 3$.

(2) VaR Baseado no Modelo Normal.

Nesse caso, assumimos ser a distribuição F uma distribuição Normal com média μ e variância σ^2 , isto é,

$$r_t \sim N(\mu, \sigma^2).$$

Podemos estimar os parâmetros usando os estimadores de máxima verossimilhança, $\bar{r} = \frac{1}{T} \sum_{t=1}^T r_t$ e $S^2 = \frac{1}{T} \sum_{t=1}^T (r_t - \bar{r})^2$. O VaR_α é, então, calculado como

$$\text{VaR}_\alpha = \bar{r} + z_\alpha S,$$

onde $S = \sqrt{S^2}$ e z_α é o quantil de probabilidade de excedência α de uma Normal padrão.

(3) VaR Baseado no Modelo Normal Assimétrico.

Nesse caso, assumimos ser a distribuição F uma distribuição Normal assimétrica com média μ e variância σ^2 e parâmetro de assimetria λ , ver definições no Capítulo 2. Sejam $\hat{\mu}$, $\hat{\lambda}$ e $\hat{\sigma}$ estimativas para μ , λ e σ , e seja $z_{\alpha, \hat{\lambda}} = \hat{F}^{-1}(1 - \alpha)$, onde \hat{F} é a Normal assimétrica *padrão* estimada. Então

$$\text{VaR}_\alpha = \hat{\mu} + z_{\alpha, \hat{\lambda}} \hat{\sigma}.$$

(4) VaR Baseado no Modelo t-student.

Nesse caso, assumimos ser a distribuição F uma distribuição t-student com ν graus de liberdade, denotada por t_ν . Notemos que a distribuição t_ν tem média zero e variância $\frac{\nu}{(\nu-2)}$, e que os quantis Q_α calculados pela maioria dos pacotes é o quantil baseado nesta distribuição. Assim, se quisermos o quantil associado à uma distribuição com variância 1, a qual denotaremos por t_ν^* , devemos dividir Q_α por $\sqrt{\nu/(\nu-2)}$. Ou seja,

$$\begin{aligned} \alpha = P(t_\nu > Q_\alpha) &= P\left(\frac{t_\nu}{\sqrt{\nu/(\nu-2)}} > \frac{Q_\alpha}{\sqrt{\nu/(\nu-2)}}\right) \\ &= P\left(t_\nu^* > \frac{Q_\alpha}{\sqrt{\nu/(\nu-2)}}\right) \\ &= P(t_\nu^* > Q_\alpha^*) \quad , \quad \nu > 2. \end{aligned}$$

Por exemplo, a t-student implementada no SPlus e no R possuem variância $\nu/(\nu-2)$. Assim o VaR_α é dado por

$$\text{VaR}_\alpha = \hat{\mu} + \frac{Q_\alpha(\hat{\nu}) \hat{\sigma}}{\sqrt{\frac{\hat{\nu}}{\hat{\nu}-2}}} = \hat{\mu} + Q_\alpha^* \hat{\sigma}$$

onde $\hat{\mu}$ e $\hat{\sigma}$ podem ser quaisquer estimadores para μ e σ , podendo ser \bar{r} e S , ou os EMV. O parâmetro ν deve ser estimado por algum método, inclusive o método da máxima verossimilhança o qual pode ser empregado para obter simultaneamente as estimativas dos três parâmetros.

(5) VaR Baseado no Modelo t-student Assimétrico.

Nesse caso, assumimos ser a distribuição F uma distribuição t_ν assimétrica com parâmetros μ , σ , e parâmetro de assimetria λ , ver várias opções no Capítulo 2. Novamente sejam $\hat{\mu}$, $\hat{\lambda}$ e $\hat{\sigma}$ estimativas para μ , λ e σ , e seja $q_{\alpha, \hat{\lambda}} = \hat{F}^{-1}(1 - \alpha)$, onde \hat{F} é a t assimétrica *padrão* estimada. Então

$$\text{VaR}_\alpha = \hat{\mu} + q_{\alpha, \hat{\lambda}} \hat{\sigma}.$$

(6) VaR Baseado no Modelo GEV.

Denotemos por $G_{\xi,\mu,\sigma}$ a função de distribuição limite para os máximos, a GEV dada por (2.18). Lembremos que o objetivo final é estimar quantis de F . Para isso, primeiro estimamos $G_{\xi,\mu,\sigma}$ utilizando os máximos coletados em blocos de tamanho n , para depois obter F a partir de

$$G_{\xi,\mu,\sigma}(x) = (F(x))^n. \quad (7.1)$$

Como o VaR_α é o quantil de probabilidade $(1 - \alpha)$ da distribuição F , reescrevemos a expressão acima como

$$G_{\xi,\mu,\sigma}(\text{VaR}_\alpha) = (1 - \alpha)^n,$$

obtendo

$$\text{VaR}_\alpha = G_{\xi,\mu,\sigma}^{-1}((1 - \alpha)^n), \quad (7.2)$$

e assim

$$\text{VaR}_\alpha = G_{\xi,\mu,\sigma}^{-1}((1 - \alpha)^n) = \begin{cases} \mu - \frac{\sigma}{\xi}(1 - p^{-\xi}), & \text{para } \xi \neq 0 \\ \mu - \sigma \log p, & \text{para } \xi = 0, \end{cases} \quad (7.3)$$

e onde $p = -\log(1 - (1 - \alpha)^n)$. Notemos que o VaR_α corresponde a um quantil de probabilidade acumulada muito menor (ou de excedência muito maior) na distribuição dos máximos. As estimativas dos parâmetros da GEV podem ser obtidas pelo método da máxima verossimilhança ou pelo método dos L-momentos.

Os praticantes do mercado preferem muitas vezes avaliar o risco através do “evento de t -períodos”. O evento de t -períodos é um conceito muito simples, que consiste basicamente na associação de um quantil da distribuição dos máximos com o número médio de períodos (blocos de tamanho n) entre suas ocorrências. Sua importância advém do fato de ser uma medida que pode ser facilmente entendida pelos leigos, correspondendo ao evento que espera-se que ocorra em média uma vez a cada t períodos.

O evento de t -períodos é calculado usando a inversa da função de distribuição da GEV, dado em (2.28)

$$G_{\xi,\mu,\sigma}^{-1}(1 - \frac{1}{t}) = \begin{cases} \mu - \frac{\sigma}{\xi}(1 - y^{-\xi}), & \text{para } \xi \neq 0 \\ \mu - \sigma \log y, & \text{para } \xi = 0, \end{cases} \quad (7.4)$$

onde $y = -\log(1 - \frac{1}{t})$.

A seguir, apresentamos a relação do VaR_α com o evento de t -períodos:

$$\begin{aligned} G_{\xi,\mu,\sigma}(\text{VaR}_\alpha) &= (F(\text{VaR}_\alpha))^n = (1 - \alpha)^n = 1 - \frac{1}{t}, \\ \text{ou } \text{VaR}_\alpha &= G_{\xi,\mu,\sigma}^{-1}((1 - \alpha)^n) = G_{\xi,\mu,\sigma}^{-1}(1 - \frac{1}{t}), \end{aligned} \quad (7.5)$$

obtendo então t a partir de

$$t = \frac{1}{1 - (1 - \alpha)^n}.$$

Como exercício sugerimos o leitor fazer um gráfico dos eventos de t -períodos, para vários valores de t , com os seus intervalos de confiança (IC), e ao lado outro gráfico com o VaR_α correspondentes com seus ICs.

(7) VaR Baseado no Modelo GEV e no Índice Extremal θ .

Esta extensão dos resultados da TVE é para séries estacionárias onde assumimos que as observações possam ser dependentes, isto é, que possa existir dependência entre valores extremos. Um resultado da TVE estabelece que a frequência e o tamanho dos conglomerados de valores extremos em uma série são características capturadas pelo seu índice extremal θ . Assim, incorporando θ à modelagem resulta em estimativas melhores para as caudas extremas e, conseqüentemente, para os quantis associados à probabilidades muito pequenas. Veja mais detalhes em Leadbetter, M., Lindgren, G. e Rootzén, H. (1983).

Dois estimadores para o índice extremal θ foram dados no Capítulo 2

$$\hat{\theta} = \frac{K_u}{N_u}.$$

e

$$\hat{\theta} = \frac{1}{m} \frac{\log \left(1 - \frac{K_u}{m}\right)}{\log \left(1 - \frac{N_u}{T}\right)}.$$

O limiar u pode ser escolhido de modo que $\frac{N_u}{T} * 100\%$ represente um certo percentual dos dados. Uma possibilidade é usar como estimativa final de θ a média das estimativas obtidas ao se fixar esse percentual em 2%, 3%, 4%, 5% e 6%.

Assim, incorporando o índice extremal θ ao modelo GEV para o cálculo do VaR, temos que

$$\text{VaR}_\alpha = G_{\xi, \hat{\mu}, \hat{\sigma}}^{-1}((1 - \alpha)^{n\hat{\theta}}). \quad (7.6)$$

Evidentemente, para séries independentes $\theta = 1$.

(8) VaR Baseado no Modelo r -maiores.

A metodologia para o cálculo do VaR usando o modelo r -maiores é idêntica à descrita para o caso GEV. A única diferença é que na estimação dos parâmetros ξ, μ, σ usamos as r -maiores estatísticas de ordem, ao invés de apenas os máximos, o que geralmente reduz bastante a variabilidade das estimativas.

Para o cálculo do VaR baseado no modelo r -maiores, primeiro extraímos as r -maiores estatísticas de ordem dos blocos de tamanho n . No caso das r -menores basta pré-multiplicar a série por (-1). Podemos utilizar blocos mensais ($n = 22$) e, no exemplo dado, fixamos $r = 3$ e ajustamos por máxima verossimilhança a densidade conjunta dada em (2.30). Em seguida obtemos os quantis de interesse da GEV e as estimativas do VaR 5%, 1% e 0,1% utilizando a fórmula (7.3).

(9) VaR Baseado no Modelo r -maiores e o Índice Extremal θ .

O método para o cálculo do VaR usando o modelo r -maiores corrigido pelo índice extremal θ é idêntico ao caso dos máximos, diferindo apenas na estimação dos parâmetros conforme explicado

acima. A incorporação de θ geralmente resulta em estimativas ainda mais acuradas das caudas da distribuição e, por conseguinte, para os quantis associados a probabilidades muito pequenas. Use fórmula (7.6).

(10) VaR Baseado no Modelo GPD.

O primeiro problema a ser resolvido é a escolha do limiar (alto) u . Isso deve ser feito de acordo com investigações empíricas e análise de gráficos. Uma possibilidade é usar a função Média dos Excessos empírica, clássica e robusta, e procurar pelo valor de u a partir do qual tal função se torna aproximadamente linear, veja detalhes no Capítulo 2.

A distribuição condicional do excesso $Y = X - u$ é denotada por F_u e no Capítulo 2 obtivemos que

$$\bar{F}(u + y) = \bar{F}_u(y) \bar{F}(u),$$

onde y representa o excesso e u o limiar. A cauda \bar{F}_u é aproximada pela GPD.

Estimamos $\bar{F}(u)$ empiricamente usando $\frac{N_u}{T}$, onde N_u é o número de r_i 's que ultrapassaram o limiar u , ou seja, o número de excedentes, e T é o tamanho da série. Note aqui o *trade-off*: Para estimar $\bar{F}(u)$ gostaríamos de um u relativamente baixo, ao passo que para estimar a GPD precisamos de um limiar alto.

Assim, sendo $u + y = \text{VaR}_\alpha$, temos

$$\bar{F}(\text{VaR}_\alpha) = \alpha = \bar{P}_{\xi, \psi}(y) p^*, \quad (7.7)$$

onde $p^* = \frac{N_u}{T}$. Calculamos, então, o quantil y de $\bar{P}_{\xi, \psi}$ tal que $\bar{P}_{\xi, \psi}(y) = \frac{\alpha}{p^*}$. Logo,

$$y = P_{\xi, \psi}^{-1} \left(\frac{\alpha}{p^*} \right),$$

e portanto temos

$$\text{VaR}_\alpha = u + P_{\xi, \psi}^{-1} \left(\frac{\alpha}{p^*} \right),$$

ou

$$\text{VaR}_\alpha = u + \frac{\psi}{\xi} \left(\left(\frac{\alpha}{p^*} \right)^{-\xi} - 1 \right),$$

onde as estimativas de (ψ, ξ) podem ser obtidas por máxima verossimilhança.

(11) VaR Baseado no Modelo GPD e Índice Extremal θ .

O objetivo aqui é lidar com o problema da dependência entre extremos no modelo de excessos além de um limiar. Esta técnica é conhecida como declusterização e corresponde a uma filtragem das observações dependentes, a fim de obter um conjunto de excessos que sejam aproximadamente independentes. Isso é feito seguindo os passos:

1. usando uma regra empírica para definir os *clusters*;

2. identificando o excesso máximo dentro de cada *cluster*; e
3. ajustando uma GPD aos máximos dos *clusters*.

O método é simples, mas tem suas limitações. Os resultados são sensíveis à especificação arbitrária dos *clusters*, e há um desperdício considerável de informação ao se descartar todas as informações exceto os máximos dos *clusters*.

Uma forma de determinar os *clusters* de extremos é especificar um limiar u e tomar os excessos de u consecutivos como pertencentes ao mesmo *cluster*. É comum considerar que um *cluster* esteja ativo até que ocorram l valores consecutivos abaixo do limiar, para algum l pré-especificado. Em seguida executamos os passos 2 e 3.

Seja N_u o número de observações excedendo o limiar u e N_c o número de *clusters* acima de u . O índice extremal pode ser estimado por

$$\hat{\theta}_u = \frac{N_c}{N_u}.$$

Assim, temos

$$\begin{aligned} \text{VaR}_\alpha &= u + P_{\xi, \psi}^{-1} \left(\frac{\alpha}{p^* \theta_u} \right), \\ \text{VaR}_\alpha &= u + \frac{\psi}{\xi} \left(\left(\frac{\alpha}{p^* \theta_u} \right)^{-\xi} - 1 \right). \end{aligned}$$

Para o exemplo que damos nesta seção, para o cálculo do VaR baseado no modelo GPD corrigido pelo índice extremal θ , consideramos para a estimação de θ que um *cluster* de excessos termina quando ocorrem duas observações consecutivas abaixo do limiar u e calculamos o VaR 5%, 1% e 0.1% dos ganhos.

Examinando a Tabela 7.1, notamos que as estimativas que levam em consideração a informação dada pelo índice extremal (GEV- θ , r -maiores- θ e GPD- θ) parecem ser adequadas para épocas de crise, já que este índice reflete justamente a presença de conglomerados de valores extremos o que aumenta o peso da cauda. De fato, o procedimento i.i.d. subestima a probabilidade de ocorrência de eventos extremos, mas seja qual for a metodologia, deveríamos sempre fazer “back-tests” para se verificar qual procedimento seria mais acurado em cada aplicação. Notemos que o procedimento ideal varia de acordo com a série utilizada, com o nível de risco e até com os objetivos da estimação.

(12) VaR Não Condicional Utilizando Cópuas

Ao se estimar o VaR de uma carteira seria desejável levar em consideração a estrutura de dependência entre seus d componentes. Neste caso uma distribuição d -variada deveria ser proposta e (estática ou dinamicamente) estimada. Já vimos que nenhuma das distribuições multivariadas conhecidas são capazes de lidar com as formas lineares e não lineares de dependência existentes entre os ativos, e nem de permitir um ajuste específico para cada marginal.

Nesta sub-seção calcularemos o VaR baseado em uma cópula estática. Podemos até ajustar modelos univariados dinâmicos às margens, mas a cópula será única, invariante no tempo. O caso da cópula condicional será visto mais a frente ao falarmos do VaR condicional.

No primeiro passo ajustamos modelos estáticos (Capítulo 2) ou dinâmicos (capítulos 4 e 5) às séries de log-retornos. Em seguida transformamos as séries de resíduos em séries de valores $Uniforme(0, 1)$ através da transformação integral da probabilidade (use a f.d.a.). Esta matriz $T \times d$ de dados transformados é utilizada para estimar os parâmetros da cópula. Várias famílias devem ser consideradas e testes podem ser usados para se definir qual delas proporciona a melhor aderência aos dados.

O cálculo do VaR_α não condicional é então feito através de simulações. Simulamos um número grande B de valores d -dimensionais da cópula ajustada, obtendo assim uma matriz $B \times d$ de valores que seriam as transformações dos resíduos. Aplicamos a inversa da f.d.a. de cada coluna, obtendo simulações dos *resíduos*. A partir desses *resíduos* e da inversa de cada modelo ajustado à cada margem obtemos finalmente a matriz de *retornos*. No caso de uma carteira, a partir desta última matriz obtemos B valores simulados da carteira e calculamos o VaR empírico (ou qualquer outra medida de risco) como o quantil de probabilidade de excedência $\alpha\%$.

(13) VaR Não Condicional Baseado em Simulações

O VaR não condicional baseado em simulações pode ser obtido para qualquer um dos métodos *paramétricos* (2) a (11). A metodologia (12) já usou simulações.

Tendo estimado a distribuição subjacente dos dados, F , simulamos um número muito grande B de replicações da série de retornos original de tamanho T . Para cada uma dessas séries simuladas calculamos o VaR empírico. No final temos uma *amostra* de B valores para o VaR_α . O VaR final pode ser a média ou a mediana dos B valores. Esta metodologia permite obter o erro padrão da medida de risco e também seu intervalo de confiança a partir dos quantis empíricos. Por exemplo, o IC de 95% seria formado pelos quantis empíricos de 2.5% e 97.5%.

7.3 VaR Condicional

Conforme já vimos, a série de log-retornos e a de seus quadrados em geral apresentam autocorrelações significativas. Os modelos condicionais estudados nos capítulos 4 e 5 baseiam-se neste fato e possibilitam fazer previsões alguns passos a frente que refletem a média e a volatilidade corrente. Em outras palavras, medidas de risco calculadas a partir de modelos condicionais refletem a situação corrente.

As metodologias condicionais variam desde um simples ajuste ARMA-GARCH seguido de cálculo do VaR a partir da distribuição preditiva estimada (um ou vários passos a frente), até modelos mais complexos utilizando ajustes ARMA-GARCH com memória longa nas séries dos ativos componentes de uma carteira e a estimação de cópulas dinâmicas com ou sem memória longa para capturar a estrutura de dependência variando no tempo.

Uma outra opção que une modelos condicionais ARMA-GARCH e modelos da TVE é também muito efetiva. Existem inúmeras publicações utilizando este tipo de procedimento, por exemplo Frey, R. e McNeil, A. J. (1998). Em Mendes, B. V. M. (2000b) uma técnica similar é utilizada onde modelos GARCH seguidos da modelagem GPD nos excessos dos resíduos são estimados de forma robusta fornecendo medidas de risco robustas. As metodologias utilizadas produzem estimativas mais precisas que aquelas obtidas pela *Riskmetrics*, que se baseiam no GARCH integrado e na suposição de normalidade, as quais, em geral, apresentam bons resultados apenas para níveis de risco maiores,

como o de 5%.

Também buscando uma forma de neutralizar os conglomerados de volatilidade das séries financeiras, Engle, R. F. e Manganelli, S. (2004) propuseram uma especificação quantílica autoregressiva condicional, a qual denominaram VaR autorregressivo condicional (CAViaR). Essa metodologia permite a inclusão de variáveis relevantes, bem como a utilização de diversas formas funcionais.

Em resumo, os métodos para o cálculo do VaR Condicional incluem:

1. VaR Condicional Abordagem do RiskMetrics
2. VaR Condicional Abordagem Econométrica: ARFIMA-FIEGARCH
3. VaR Condicional Abordagem Econométrica: ARFIMA-FIEGARCH e GPD
4. VaR Condicional Abordagem Econométrica: CAVIAR
5. VaR Condicional Abordagem Econométrica: ARFIMA-FIEGARCH - Cópias Dinâmicas

Selecionamos as seguintes possibilidades:

(1) Abordagem do RiskMetrics.

J. P. Morgan desenvolveu a metodologia RiskMetricsTM para o cálculo do VaR condicional. Na sua forma mais simples essa metodologia assume que

$$r_t | \mathcal{I}_{t-1} \sim N(\mu_t, \sigma_t^2)$$

onde r_t é o log-retorno diário, e onde a média e a variância condicionais evoluem no tempo, de acordo com

$$\mu_t = 0 \quad \sigma_t^2 = \alpha_0 + \beta \sigma_{t-1}^2 + (1 - \beta) r_{t-1}^2 \quad 0 < \beta < 1,$$

isto é, um IGARCH(1,1).

Veremos a seguir as vantagens dessa modelagem. Seja t a origem e k o horizonte. Então o log-retorno de k -períodos é:

$$r_t[k] = r_{t+1} + \dots + r_{t+k-1} + r_{t+k}.$$

Como $r_t = p_t - p_{t-1} = a_t$, temos $r_t[k] = a_{t+1} + \dots + a_{t+k}$. Também temos $a_t = \sigma_t \epsilon_t$ e $\text{var}(r_t[k]) = \sum_{i=1}^k \text{var}(a_{t+i} | \mathcal{I}_t)$, onde as variâncias $\text{var}(a_{t+i} | \mathcal{I}_t) = E[\sigma_{t+i}^2 | \mathcal{I}_t]$ são obtidas recursivamente.

Considere o modelo IGARCH(1,1)

$$\sigma_t^2 = \beta \sigma_{t-1}^2 + (1 - \beta) r_{t-1}^2.$$

Sabendo-se que $r_{t-1} = a_{t-1} = \sigma_{t-1} \epsilon_{t-1}$, obtemos

$$\sigma_t^2 = \beta \sigma_{t-1}^2 + (1 - \beta) \sigma_{t-1}^2 \epsilon_{t-1}^2$$

$$\begin{aligned}
\sigma_t^2 &= \beta\sigma_{t-1}^2 + \sigma_{t-1}^2\epsilon_{t-1}^2 + \sigma_{t-1}^2 - \beta\sigma_{t-1}^2\epsilon_{t-1}^2 - \sigma_{t-1}^2 \\
\sigma_t^2 &= \sigma_{t-1}^2 + \sigma_{t-1}^2(\epsilon_{t-1}^2 - 1) - \beta\sigma_{t-1}^2(\epsilon_{t-1}^2 - 1) \\
\sigma_t^2 &= \sigma_{t-1}^2 + (1 - \beta)\sigma_{t-1}^2(\epsilon_{t-1}^2 - 1).
\end{aligned}$$

A última expressão naturalmente também vale para todo t ou $t + i$, e como $E[\epsilon_{t-1}^2 - 1] = 0$, temos $E[\sigma_{t+i}^2|\mathcal{I}_t] = E[\sigma_{t+i-1}^2|\mathcal{I}_t]$, para $i = 2, \dots, k$. Então,

$$\sigma_t^2[k] = k\sigma_{t+1}^2.$$

Resumindo, $r_t[k]|\mathcal{I}_t \sim N(0, k\sigma_{t+1}^2)$, com a variância condicional seguindo um modelo $IGARCH(1, 1)$. Assim, para o cálculo do VaR_α , e considerando as suposições da metodologia RiskMetrics, devemos ajustar um $IGARCH(1, 1)$ com distribuição condicional normal e estimar a volatilidade um passo a frente. Isto é, estando no tempo t , obter $\widehat{\sigma_{t+1}^2}$ e calcular o VaR_α k dias a frente como

$$\text{VaR}_{t,\alpha}[k] = z_\alpha \sqrt{k} \widehat{\sigma_{t+1}},$$

onde z_α é o quantil de probabilidade de excedência α . Esse método é fácil de implementar, porém baseia-se na suposição de normalidade, o que pode subestimar o VaR. Além disso, a *regra da raiz quadrada de k* somente vale se a média de r_t for zero e o modelo para a volatilidade for o $IGARCH(1, 1)$.

A fórmula dada vale para o cálculo do VaR de apenas uma posição, o qual é calculado no caso de um-passo-a-frente como

$$\text{VaR}_{t,\alpha} = (\text{Valor da Posição}) z_\alpha \widehat{\sigma_{t+1}}.$$

Suponha agora que r_1 e r_2 representem retornos de 2 posições ou carteiras, e seja ρ_{12} o coeficiente de correlação linear entre esses retornos. Isto é, $\rho_{12} = \frac{\text{cov}(r_1, r_2)}{\sqrt{\text{var}(r_1)\text{var}(r_2)}}$. Para o cálculo do VaR_α de *Múltiplas Posições* o RiskMetrics adota a seguinte fórmula (facilmente deduzível)

$$\text{VaR}_\alpha = \sqrt{(\text{VaR}_\alpha^1)^2 + (\text{VaR}_\alpha^2)^2 + 2\rho_{12}\text{VaR}_\alpha^1\text{VaR}_\alpha^2},$$

onde VaR_α^i representa o VaR de risco α do ativo i , $i = 1, 2$.

(2) Abordagem Econométrica: ARFIMA-FIEGARCH.

Calculamos o VaR_α condicional a partir do ajuste do modelo ARFIMA (p, d, q) -FIGARCH (m, d, s) . Considere as equações para média e volatilidade condicionais

$$\begin{aligned}
r_t &= c + \sum_{i=1}^p \phi_i r_{t-i} + a_t - \sum_{j=1}^q \theta_j a_{t-j}, \\
a_t &= \sigma_t \epsilon_t, \\
\sigma_t^2 &= \alpha_0 + \sum_{i=1}^m \alpha_i a_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2,
\end{aligned}$$

onde $\epsilon_t \text{ iid} \sim F(0, 1)$ com F sendo a distribuição normal ou a t-student com ν graus de liberdade. Nas equações acima podemos ainda incluir a modelagem da memória longa tanto na média como na variância, e ainda o efeito da volatilidade na média (GARCH-in-Mean). Iremos, ainda, acrescentar um modelo da TVE para as caudas das inovações.

Se escolhermos F como sendo a $N(0, 1)$, a distribuição preditiva um passo a frente será a

$$N(\widehat{r}_t(1), \widehat{\sigma}_t^2(1))$$

onde $\widehat{r}_t(1)$ e $\widehat{\sigma}_t^2(1)$ são as previsões um passo a frente da média e da variância obtidas a partir do ajuste do modelo dado pelas equações acima. Para calcular o VaR com probabilidade de excedência α calculamos

$$\text{VaR}_\alpha = \widehat{r}_t(1) + z_\alpha \widehat{\sigma}_t(1)$$

já que a probabilidade da variável aleatória normal padrão $\frac{r_{t+1} - \widehat{r}_t(1)}{\widehat{\sigma}_t(1)}$ ser maior ou igual a z_α é igual a α .

Para o cálculo do VaR k períodos a frente, é importante lembrar que a *regra da raiz quadrada* não se aplica. Precisamos saber especificar a distribuição preditiva de $r_t[k]$, o retorno k dias a frente, $r_t[k] = r_{t+1} + \dots + r_{t+k}$. Seja o modelo econométrico dado pelas equações acima. Para a previsão da média, utilizamos a esperança condicional, a qual minimiza a esperança da perda quadrática. Para os modelos ARMA, $\widehat{r}_t[k] = r_t(1) + \dots + r_t(k)$, onde $r_t(l)$ representa a previsão l passos a frente.

Precisamos agora obter o erro de previsão. Para isso, usaremos a representação $\text{MA}(\infty)$ para o $\text{ARMA}(p, q)$ (para os modelos ARFIMA veja Tsay, R. S. (2005)). Nesse caso,

$$e_t(l) = r_{t+l} - r_t(l)$$

e, portanto,

$$\begin{aligned} e_t(l) = & \mu + a_{t+l} + \psi_1 a_{t+l-1} + \psi_2 a_{t+l-2} + \dots + \psi_{l-1} a_{t+l-l+1} + \psi_{l-2} a_t + \psi_{l-3} a_{t-1} + \psi_{l-4} a_{t-2} + \dots \\ & + \dots - (\mu + \text{todos os termos a partir de } t \text{ para trás}). \end{aligned}$$

Então,

$$e_t(l) = a_{t+l} + \psi_1 a_{t+l-1} + \dots + \psi_{l-1} a_{t+1}. \quad (*)$$

Sabendo que as previsões são $\widehat{r}_t[k] = r_t(1) + \dots + r_t(k)$, o erro de previsão de $\widehat{r}_t[k]$ é a soma dos erros de previsão de 1-passo

$$e_t[k] = e_t(1) + e_t(2) + \dots + e_t(k)$$

e usando (*) obtemos

$$e_t[k] = a_{t+1} + (a_{t+2} + \psi_1 a_{t+1}) + (\dots) + \sum_{i=0}^{k-1} \psi_i a_{t+k-i}.$$

Isto é,

$$e_t[k] = a_{t+k} + (1 + \psi_1) a_{t+k-1} + \dots + \left(\sum_{i=0}^{k-1} \psi_i \right) a_{t+1}, \quad (**)$$

uma soma ponderada de choques a partir de $(t+1)$ até $(t+k)$.

Derivamos agora a previsão da volatilidade de $r_t[k]$, a variância preditiva k períodos a frente, $var(e_t[k]|\mathcal{I}_t)$.

Sabendo que $a_{t+i} = \sigma_{t+i}\epsilon_{t+i}$ e $\epsilon_t \sim i.i.d.$, e usando estas suposições em $(**)$ obtemos

$$var(e_t[k]|\mathcal{I}_t) = \sigma_t^2(k) + (1 + \psi_1)^2 \sigma_t^2(k-1) + \cdots + \left(\sum_{i=0}^{k-1} \psi_i\right)^2 \sigma_t^2(1). \quad (***)$$

As previsões da volatilidade a partir da origem t , $\sigma_t^2(l)$, são obtidas usando os modelos GARCH, as quais são inseridas em $(***)$. Temos então a distribuição preditiva do retorno de k -períodos, $r_t[k]$, dada pela F assumida e sua média e variância obtidas acima. Novamente recomendamos a especificação t -student.

(3) Abordagem Econométrica: ARFIMA-FIEGARCH e GPD.

Suponha que os log-retornos diários, denotados por r_1, \dots, r_T , sejam realizações de um processo estritamente estacionário, onde a média e a variância de r_t variam no tempo, de acordo com

$$r_t = \mu_t + \sigma_t Z_t, \quad (7.8)$$

e que as inovações Z_t sejam v.a.s i.i.d. com distribuição $F_Z(\cdot)$, com média zero e variância unitária.

Em geral, um modelo AR(1) combinado com um GARCH(1,1) é adequado e suficiente para modelar a média e a volatilidade condicional dos retornos financeiros. Isto é,

$$r_t = \phi r_{t-1} + a_t, \quad (7.9)$$

onde as inovações a_t seguem um modelo GARCH(1,1). Assim, condicionalmente à toda informação passada até o tempo t , I_{t-1} , o quadrado da volatilidade no tempo t é dado por

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \quad (7.10)$$

onde $\alpha_0 > 0$, $\alpha_1 \geq 0$, e $\alpha_1 + \beta_1 < 1$.

Para modelar as caudas da distribuição dos resíduos padronizados do ajuste AR(1)+GARCH(1,1)

$$Z_1, \dots, Z_T = \frac{r_1 - \mu_1}{\sigma_1}, \dots, \frac{r_T - \mu_T}{\sigma_T} \quad (7.11)$$

usamos uma GPD aplicada aos excessos além de um limiar alto u .

Para o cálculo do VaR, seja N_u o número de resíduos Z_t , $t = 1, \dots, T$, excedendo um limiar alto u , e sejam Y_1, \dots, Y_{N_u} esses excessos. Recordemos que

$$\bar{F}(u+y) = \bar{F}(u) \cdot \bar{F}_u(y).$$

Essa expressão será usada para calcular probabilidades na cauda de r_t ($\bar{F}(u+y)$) para quantis muito extremos, usando o estimador empírico N_u/T para $\bar{F}(u)$ e aproximando $\bar{F}_u(y)$ pela GPD. Assim,

temos que

$$\bar{F}_Z(u+y) = \frac{N_u}{T} \left(1 + \xi \frac{y}{\psi}\right)^{-1/\xi}, \quad (7.12)$$

para $y \geq 0$ quando $\xi \geq 0$. Usando a inversa $\bar{F}_Z^{-1}(\cdot)$ em (7.12) obtemos um estimador do quantil

$$z_\alpha = u + \frac{\psi}{\xi} \left(\left(\frac{\alpha}{N_u/T} \right)^{-\xi} - 1 \right). \quad (7.13)$$

Isto é, z_α é o $(1-\alpha)$ -quantil da distribuição de Z_t , um valor com probabilidade de excedência α .

O VaR condicional um passo a frente é

$$\text{VaR}_\alpha = \mu_{t+1} + \sigma_{t+1} z_\alpha, \quad (7.14)$$

para $0 < \alpha < 1$ e $\alpha < \frac{N_u}{T}$, onde z_α não depende de t , e é estimado usando (7.13).

Conforme já pudemos observar, ao utilizarmos a GPD, a escolha do limiar u é uma decisão crítica. Para os resíduos ordenados $z_{(1)} \geq z_{(2)} \geq \dots \geq z_{(T)}$, pode-se escolher o limiar u como o $(r+1)$ -ésimo maior resíduo $z_{(r+1)}$ tal que o número de excedências N_u represente um certo percentual dos dados. Então, ajusta-se a $P_{\xi, \psi}$ (GPD) aos r excessos $(z_{(1)} - z_{(r+1)}, \dots, z_{(r)} - z_{(r+1)})$. Assim, para $\alpha < r/T$, o estimador de cauda (7.12) e o estimador do quantil (7.13) são calculados usando

$$\hat{\bar{F}}_Z(z) = \frac{r}{T} \left(1 + \hat{\xi}_r \frac{z - z_{(r+1)}}{\hat{\psi}_r} \right)^{-\frac{1}{\hat{\xi}_r}}, \quad (7.15)$$

e

$$\hat{z}_\alpha = z_{(r+1)} + \frac{\hat{\psi}_r}{\hat{\xi}_r} \left(\left(\frac{\alpha}{r/T} \right)^{-\hat{\xi}_r} - 1 \right). \quad (7.16)$$

Para as perdas do índice do Brasil as estimativas do VaR condicional são dadas na Tabela 7.2.

Tabela 7.2: Estimativas do VaR condicional para $\alpha = 0.001, 0.01$, e 0.05 e para as perdas do índice brasileiro.

Metodologia	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.001$
1.VaR Condicional Baseado no GARCH(1,1) Normal	-2.8879	-4.1250	-5.5117
2.VaR Condicional Baseado no GARCH(1,1) t-student(10)	-3.0187	-4.6031	-6.9014
3.VaR Condicional Baseado no GARCH(1,1) Normal-GPD	-3.0493	-4.8424	-7.0523

(3) VaR Autorregressivo Condicional

Engle, R. F. e Manganelli, S. (2004) desenvolveram a metodologia para o cálculo do VaR autorregressivo condicional (CAViaR). A forma geral da especificação CAViaR é

$$\text{VaR}_t = \beta_0 + \sum_{i=1}^p \beta_i \text{VaR}_{t-i} + g(\beta_{p+1}, \dots, \beta_{p+q}; \mathcal{F}_{t-1}), \quad (7.17)$$

onde \mathcal{F}_{t-1} são as informações passadas.

Na maioria dos casos práticos, a especificação dada em (7.17) se reduz a um modelo de primeira ordem do tipo

$$\text{VaR}_t = \beta_0 + \beta_1 \text{VaR}_{t-1} + g(\beta_2, r_{t-1}, \text{VaR}_{t-1}). \quad (7.18)$$

O termo autorregressivo de primeira ordem assegura que o VaR varia suavemente no tempo, enquanto $g(\beta_2, r_{t-1}, \text{VaR}_{t-1})$ faz a ligação de VaR_t a r_{t-1} , ou seja, $g(\cdot)$ mede quanto o VaR pode variar baseado na informação trazida pelo retorno. Notemos que, para o processo (7.17) não ser explosivo, as raízes de

$$1 - \beta_1 z - \beta_2 z^2 - \dots - \beta_p z^p = 0$$

devem estar fora do círculo unitário.

Algumas especificações CAViaR são:

(3i) Absoluta simétrica:

$$\text{VaR}_t = \beta_0 + \beta_1 \text{VaR}_{t-1} + \beta_2 |r_t - 1|. \quad (7.19)$$

Essa especificação responde simetricamente aos retornos passados, visto que $g(\cdot)$ é a função módulo.

(3ii) Inclinada assimétrica:

$$\text{VaR}_t = \beta_0 + \beta_1 \text{VaR}_{t-1} + \beta_2 \max(r_{t-1}, 0) + \beta_3 \min(r_{t-1}, 0). \quad (7.20)$$

Essa especificação responde diferentemente aos retornos passados positivos e negativos, a menos que $\beta_2 = \beta_3$.

(3iii) GARCH(1,1) indireto:

$$\text{VaR}_t = \sqrt{\beta_0 + \beta_1 \text{VaR}_{t-1}^2 + \beta_2 r_{t-1}^2}. \quad (7.21)$$

Assim como a especificação absoluta simétrica, essa especificação responde aos retornos passados de forma simétrica, visto que $g(\cdot)$ é quadrática.

Esses modelos retratam casos bastante específicos. As especificações CAViaR são mais gerais, podendo ser utilizadas em situações em que a volatilidade é constante, mas a distribuição do erro é não-constante, ou quando ambas não são constantes.

(5) VaR Condicional Utilizando Cópulas Condicionais

Para capturar toda a dinâmica encontrada na distribuição multivariada dos log-retornos, muitas vezes não é suficiente ajustar apenas modelos condicionais marginais para a volatilidade. Torna-se também necessário capturar as mudanças temporais na estrutura de dependência. Isto pode ser feito combinando models GARCH (mais geralmente FIEGARCH), e cópulas variando no tempo, veja

Mendes, B. V. M. (2005b) e Van den Goorbergh, R. W. J., Genest, C. e Werker, B. J. M. (2005). Lembremos ainda que além da estimação do VaR, a estimação variando no tempo da distribuição multivariada é base para várias importantes aplicações em finanças, por exemplo, seleção dos componentes de uma carteira, apreçamento de opções, e modelos de apreçamento de ativos.

Nesta metodologia todos os modelos univariados e a cópula possuem seus parâmetros variando no tempo. O algoritmo é similar ao descrito nos itens (12) e (13) e tem como base a simulação de valores uniformes padrão a partir da cópula dinâmica estimada. As previsões são feitas a partir dos modelos condicionais estimados obtendo-se a distribuição multivariada preditiva um ou vários passos a frente. A partir da matriz com B valores simulados da carteira calculamos o VaR empírico como o quantil de probabilidade de excedência $\alpha\%$. Podemos obter o erro padrão da medida e seu intervalo de confiança.

Para verificar a acurácia de todas essas metodologias podemos aplicar o teste de Kupiec, conforme veremos na Seção 7.6.

7.4 Shortfall Condicional

Como o VaR não permite inferir a respeito da magnitude das perdas às quais se refere, foram propostas na literatura medidas alternativas de risco, tais como a Perda Média (*expected shortfall*) e a Perda Mediana (*median shortfall*).

A Perda Média (Artzner, P., Delbaen, F., Eber, J. e Heath, D., 1997) é o valor esperado $E[P|P > \text{VaR}_\alpha]$, onde a variável P representa a perda. Então, essa medida pode ser entendida como um valor médio da perda além de um certo limiar, aqui o VaR_α . Para $0 < \alpha < 1$, a Perda Média condicional um passo a frente é definida como

$$\bar{S}_{t,\alpha} = E[r_{t+1}|r_{t+1} > \text{VaR}_{t,\alpha}, I_t]. \quad (7.22)$$

A Perda Mediana condicional é uma variação robusta da Perda Média (veja Mendes, B. V. M. (2004)), que pode ser interpretada como um valor de corte tal que as perdas além do VaR_α ocorrerão 50% das vezes abaixo ou acima do mesmo. A Perda Mediana parece ser uma quantidade mais confiável quando estimamos um valor central para perdas além de um certo limiar para uma distribuição assimétrica tal como a GPD. Ela é definida como

$$\underline{S}_{t,\alpha} = \text{Mediana}[r_{t+1}|r_{t+1} > \text{VaR}_{t,\alpha}, I_t]. \quad (7.23)$$

Para calcular as Perdas Média e Mediana, notemos novamente que buscamos a distribuição preditiva do retorno, o qual é especificado no tempo t como

$$r_t = \mu_t + \sigma_t Z_t$$

sendo que os resíduos Z_t ou simplesmente Z tem distribuição F_Z (não condicional). A distribuição preditiva um passo a frente é $F_Z\left(\frac{r_{t+1}-\mu_{t+1}}{\sigma_{t+1}}\right)$ e segue que

$$\bar{S}_{t,\alpha} = \mu_{t+1} + \sigma_{t+1} E[Z|Z > z_\alpha]. \quad (7.24)$$

Similarmente, a Perda Mediana condicional é definida como

$$\underline{S}_{t,\alpha} = \mu_{t+1} + \sigma_{t+1} \text{Mediana}[Z|Z > z_\alpha] . \quad (7.25)$$

A partir dos resultados da TVE (veja detalhes em Mendes, B. V. M. (2004)) obtemos

$$E[Z|Z > z_\alpha] = z_\alpha + \frac{\psi + \xi(z_\alpha - u)}{1 - \xi} , \quad (7.26)$$

e

$$\text{Mediana}[Z|Z > z_\alpha] = z_\alpha + \frac{[2^\xi - 1][\psi + \xi(z_\alpha - u)]}{\xi} . \quad (7.27)$$

As expressões (7.26) e (7.27) serão usadas para calcular as medidas de risco (7.24) e (7.25).

A Tabela 7.3 apresenta as estimativas das duas medidas de risco condicionais (*shortfall* médio e *shortfall* mediano), calculadas para as perdas da série do índice brasileiro sob o modelo GARCH(1, 1) Normal-GPD.

Tabela 7.3: *Perda Média* (\bar{S}_t) e *Perda Mediana* (\underline{S}_t), baseados no modelo GARCH Normal-GPD.

	Estimação Clássica ($\epsilon_t \mathcal{I}_{t-1} \sim N(0, 1)$)					
	$\bar{S}_{.05}$	$\underline{S}_{.05}$	$\bar{S}_{.01}$	$\underline{S}_{.01}$	$\bar{S}_{.001}$	$\underline{S}_{.001}$
Índice Brasil	-12,75	-12,17	-15,99	-15,44	-20,43	-19,91

Analisando a Tabela 7.3, notamos que a Perda Média superestima o risco quando comparada à Perda Mediana, para todas as probabilidades, o que era esperado, já que os eventos extremos têm mais influência na estimativa da média do que na mediana.

7.5 Drawdown-em-Risco e Máximo Drawdown-em-Risco

Os clientes de instituições financeiras geralmente acompanham de perto a performance de seus investimentos, mesmo quando seus objetivos não são de curto prazo. Embora a volatilidade aumente a tensão, uma sequência de d quedas consecutivas no preço P_t de um portfolio pode fazer o investidor se retirar do mercado. Por outro lado, uma sequência de perdas de longa duração pode ser um teste extremo para um administrador de fundos e sua metodologia de negócios. Nesses períodos de turbulência, medidas de risco estáticas de um único período, como o VaR, em geral não dão uma descrição completa do risco.

De fato, séries financeiras apresentam períodos de turbulência geralmente caracterizados por quedas consecutivas nos preços denominadas *drawdowns*. Nesses casos, a percepção do risco é diferente, e para medi-lo adequadamente podemos utilizar uma medida alternativa conhecida como *Drawdown-at-Risk* (DaR_α).

O *drawdown* pode ser definido como a soma de d retornos negativos consecutivos, e assim representa a perda acumulada após uma sequência de retornos negativos. O *drawdown* e seu complementar, o

drawup, são variáveis que informam sobre a existência de uma estrutura de dependência local, e fornecem uma nova medida de risco, diferente daquelas amplamente difundidas e aceitas no mercado.

Seja P_t o preço de um investimento no dia t e r_t o retorno logarítmico diário percentual, e assumamos que $\{r_t\}_{t \geq 1}$ seja uma sequência estacionária. Seja P_{t-1} um máximo local, ou seja, $P_{t-2} \leq P_{t-1} > P_t$, e considere a sequência de $d+1$ quedas consecutivas nos preços, isto é, $P_{t-1} > P_t > P_{t+1} > \dots > P_{t+d}$. Isso resulta em uma sequência de d retornos negativos, $r_t, r_{t+1}, \dots, r_{t+d}$. O *drawdown* X com duração D igual a d dias é definido como $X = (\ln P_{t+d} - \ln P_{t-1}) * 100 = r_{t+d} + r_{t+d-1} + \dots + r_t$. O mesmo conceito se aplica ao *drawup*.

A definição acima permite modelar separadamente a duração D e a severidade X das perdas, a exemplo do modelo tradicional de risco atuarial. Considere o período fixado de T dias úteis e a coleção de *drawdowns* $X_1, \dots, X_S, S < T$, ocorridos neste período. O máximo *drawdown* é definido como $M = \min(X_1, \dots, X_S)$. Um resultado importante da TVE dado em Embrechts, Klüppelberg e Mikosch (1997) estabelece que a cauda de uma MGPD também segue uma MGPD.

O *Drawdown* em Risco com probabilidade de excedência α , DaR_α , é o $(1-\alpha)$ -quantil da distribuição G do *drawdown*, isto é, $G^{-1}(1-\alpha) = \text{DaR}_\alpha$, onde $G^{-1}(\cdot)$ é a inversa generalizada de G . A distribuição G pode ser a MGPD dada em (2.42). Essa medida quantifica as possíveis perdas cumulativas que um investimento pode alcançar em $\alpha*100\%$ das vezes durante um horizonte de tempo fixo.

Chekhlov *et al* (2003) propuseram o uso do DaR condicional (CDaR). Essa medida tem conceito similar ao do Shortfall e é dada pelo valor esperado da distribuição dos *drawdowns* que excedem um limiar, usualmente o DaR_α , para um α pré-especificado. Chekhlov *et al* (2003) propuseram estimar o limiar DaR_α empiricamente e calcular o CDaR_α como a média dos *drawdowns* que excedem este limiar. Mendes, B. V. M. e Brandi, V. R. (2004) propuseram uma estimativa paramétrica do CDaR_α , ajustando uma MGPD à cauda dos *drawdowns*, isto é, às observações excedendo o DaR_α calculado com base no modelo MGPD.

O máximo *Drawdown* é definido como o maior *drawdown* ocorrendo em um período de tempo fixo. Podemos ajustar uma MGPD aos dados de máximo *drawdown*. O Máximo *Drawdown* em Risco com probabilidade de excedência α , MDaR_α , é tão-somente o $(1-\alpha)$ -quantil da distribuição G do máximo *drawdown*, isto é, $G^{-1}(1-\alpha) = \text{MDaR}_\alpha$, onde $G^{-1}(\cdot)$ representa a inversa generalizada de G . É uma medida que quantifica a possibilidade máxima de perdas cumulativas que um investimento pode alcançar em $\alpha*100\%$ das vezes durante um horizonte de tempo fixo, e que pode ser usada para discriminar e classificar investimentos. Por exemplo, dado um conjunto de possíveis investimentos, é possível fixar um valor de risco admissível e, então, escolher o investimento que atinge o nível de aversão ao risco do investidor. Alternativamente, para qualquer risco fixado α é possível escolher o investimento que minimiza o valor do MDaR_α .

Mendes, B. V. M. e Leal, R. P. C. (2005) propuseram o MDaR_α condicional, denotado por CMDaR_α , como sendo o valor esperado da distribuição dos máximos *drawdowns* que excedem um limiar, geralmente um MDaR_α , para um α pré-especificado. O CMDaR_α mede a perda máxima acumulada esperada dado que a perda acumulada é maior que algum valor pré-determinado.

A Tabela 7.4 apresenta algumas estatísticas básicas para a série de retornos do índice brasileiro.

Essa tabela apresenta o tamanho da amostra de *drawdowns*, os três menores *drawdowns* (%), o tamanho da amostra de *drawups* e os três maiores *drawups* (%).

Tabela 7.4: *Estatísticas básicas para as severidade dos drawdowns (DD) e drawups (DU) do índice brasileiro.*

Índice	nD	3 menores DD (%)			nU	3 maiores DU (%)		
Índice Brasileiro	780	-44,85	-43,80	-37,79	790	39,87	27,05	25,14

Notação na tabela: nD : tamanho amostral dos DD; e nU : tamanho amostral dos DU.

7.6 Qual a melhor estimativa do VaR?

Diante de tantas opções para o cálculo do VaR, se faz necessário de alguma forma avaliar a performance das diferentes metodologias existentes. Como saber qual é a melhor estimativa do VaR para uma certa instituição, ou qual é a mais adequada para uma determinada situação? Existem alguns métodos para se acessar a qualidade de uma metodologia para o cálculo do VaR. Isto pode ser feito através de testes estatísticos, os back-testes, ou experimentos de simulação. Assim como ocorre com previsões, a dificuldade é a de que as conclusões atingidas valem apenas dadas todas as condições fixadas de coleta de dados, estimação dos modelos, e metodologia de cálculo do VaR.

Seja V o número de violações do VaR, isto é, o número de dias (ou períodos) em que a perda foi maior que o VaR estimado. Os teste estatísticos são em geral baseados na contagem V de violações do VaR e em alguma estatística baseada na distribuição binomial. O teste mais conhecido é o de cobertura de Kupiec, P. H. (1995), baseado no modelo binomial que compara o número de violações esperado e o observado. Sua hipótese nula é a de que a probabilidade α ($\alpha = P(r_t > \text{VaR}_\alpha)$ ou $\alpha = P(r_t < \text{VaR}_\alpha)$) está corretamente especificada, isto é, $H_0: p = \alpha$, onde p é a verdadeira probabilidade de sucesso. Sob H_0 e de acordo com a distribuição binomial, a probabilidade de se observar v violações em T tentativas, $0 \leq v \leq T$, é dada por $\binom{T}{v}(1 - \alpha)^{T-v}\alpha^v$, de tal maneira que a estatística teste é baseada na razão das log-verossimilhanças (veja Apêndice 2.9 do Capítulo 2) e tem distribuição assintótica chi-quadrado com 1 grau de liberdade. A prática nos diz que em geral este teste aceita a hipótese nula para a maioria das metodologias comparadas não conseguindo discriminar os procedimentos, mas o poder do teste (sua capacidade de rejeitar um modelo ruim) cresce com T .

Para mais detalhes sobre *backtesting* veja Ferreira, C. A. (2013), onde outros testes relacionados são estudados, por exemplo, testes de independência serial das violações; backtest baseado na *duration*; backtest para Perda Média; etc. Existem ainda outras metodologias para a verificação da performance de medidas de risco baseadas em diferentes funções de perda, veja por exemplo Christoffersen, P. e Pelletier, D. (2004).

É sempre interessante avaliar a performance das diversas metodologias através de simulações. Em geral simula-se a partir do processo gerador estimado um número grande B de séries com tamanho T igual à série em estudo e, através de uma janela deslizante, estima-se o VaR_α , para α fixo, um passo a frente (ou vários) de acordo com todas as metodologias sob comparação. Para cada série teremos uma proporção de violações do VaR, e para as B séries teremos uma distribuição da proporção de violações. Em Engle, R. F. e Manganelli, S. (2004) podemos ver esta abordagem onde os autores comparam diversas metodologias para o cálculo do VaR usando estatísticas descritivas simples do

número de violações. Podemos estudar a distribuição do número de violações também a partir de análises gráficas e análises exploratórias. Podemos ainda calcular o erro médio quadrático e o erro mediano absoluto. Todas essas informações serão usadas para decidir qual a melhor metodologia para o cálculo do VaR para um determinado processo gerador e uma determinada probabilidade de excedência α .

7.7 Exercícios

Exercício 1: Experimento tipo Monte Carlo. Gere 10000 séries de tamanho $T = 1000$ de uma distribuição t_2 . Para cada uma delas, e para três valores diferentes de u pré-especificados, calcule o VaR_α conforme a metodologia da seção 7.2 (10), e para $\alpha = 0.001$. Calcule também o VaR_α empírico. Faça um resumo contendo a média, a mediana, o desvio padrão, os quantis 1% e 99% dos resultados.

Exercício 2: Para uma série financeira repita os cálculos da seção 7.2. Isto é, calcule o VaR não condicional usando as metodologias ilustradas na Tabela 7.1.

Exercício 3: Escolha uma série de retornos financeiros diários de tamanho não inferior a 1200 observações. Ajuste o modelo adequado para calcular o VaR da seção 7.3 (3) para $\alpha = 0.01, 0.001$, para todos os dias. Faça o gráfico da série no tempo e superponha os dois VaR_α . Calcule a proporção observada de dias quando o retorno diário (negativo) foi mais extremo que o VaR_α estimado. Compare com o valor de α .

Exercício 4: Escolha uma série de retornos financeiros diários de tamanho não inferior a 1200 observações. Ajuste o modelo da para o cálculo do VaR dados em 7.3 (2) e (3) . Faça todas as análises gráficas dos resíduos z_α sugeridas nos capítulos 1 e 5. Observe que o modelo GARCH captura todos os fatos estilizados observados em séries financeiras.

Exercício 5: Faça um estudo de simulação com o objetivo de identificar e compreender o efeito da suposição (errada) de normalidade na estimação do modelo GARCH. Estimadores obtidos sob essa suposição são chamados de quasi-verossimilhança. Gere séries com estrutura GARCH e com distribuição condicional t_ν , $\nu = 2, 3, 4, 5, 6$, fixe um limiar a 5% dos dados, ajuste a GPD, e observe o valor de ξ . Notemos que para a t_ν temos $\alpha = \nu$ e $\xi = \frac{1}{\nu}$. Observe se os resíduos ficam “achados” devido a um valor possivelmente inflado de σ_t .

Índice Remissivo

- ARFIMA, 120
- ARIMA(p, d, q), 95
- Capítulo 1: Retornos Financeiros, 11
- Capítulo 2: Modelagem Univariada N.C., 29
- Capítulo 3: Modelagem Multivariada N.C., 75
- Capítulo 4: Modelagem Univariada C., 87
- Capítulo 5: Volatilidade, 133
- Capítulo 6: Modelagem Multivariada C., 165
- Capítulo 7: Medidas de Risco, 195
- Cenários de stress, 57
- Coefficiente de assimetria, 18
- Coefficiente de correlação τ de Kendall, 23
- Coefficiente de correlação linear, 38
- Coefficiente de curtose, 18
- Coefficiente de dependência de cauda, 80
- Cointegração, 168
- Cointegração com limiar, 175
- Convergência em probabilidade, 69
- Convergência quase certa, 69
- Copulas, 77
- Distribuição esférica, 36
- Distribuição estável, 34
- Distribuição t multivariada assimétrica, 42
- Distribuição t -assimétrica, 43
- Distribuição t -assimétrica de Hansen, 43
- Distribuição t -Student, 41
- Distribuição elíptica, 37
- Distribuição elíptica: estimação, 39
- Distribuição empírica, 31
- Distribuição Normal, 31
- Distribuição Normal assimétrica, 35
- Distribuições do mesmo tipo, 69
- Domínio de atração, 34, 69
- Domínio de atração da GEV, 47
- Drawdown, 212
- Ergodicidade, 87
- Escolha do limiar, 64
- Estacionariedade, 12, 87
- Estacionariedade estrita, 12
- Estimador consistente, 32
- Estimador de Hill, 66
- Evento de t períodos, 57
- Fatos estilizados, 20
- Fronteira eficiente, 25
- Função de autocorrelação parcial, 102
- Função média podada dos excessos, 71
- Função mediana dos excessos, 64
- GARCH, 141
- GEV, 52
- GPD, 60, 64
- Independência, 19
- Índice de cauda, 47
- Índice extremal, 50
- Inversa generalizada, 70
- Kupiec, 214
- Média e Variância de Markowitz, 31
- Método dos L-momentos, 51
- Método dos momentos, 70
- MGPD, 68
- Modelo HAR-RV, 158
- Modelo VECM, 169

Momentos e quantis da GPD, 70

Movimento Browniano, 92

MVM, 31

Pair-copulas, 82

Passeio aleatório, 26, 30

Passeio aleatório com drift, 26

Persistência, 142

Processo puramente aleatório, 19

Processos estocásticos, 90

Processos lineares, 93

Range realizado, 161

Retorno, 12, 14, 15

Shortfall, 211

Simetria elíptica, 43

Subaditividade do VaR, 37

Teste da razão de verossimilhanças, 72

Teste de aderência de Kolmogorov, 31

Teste de Johansen, 176

Teste de normalidade, 18

Teste de raiz unitária, 127

Teste Jarque-Bera, 18

Teste KPSS, 16

Teste KS (GOF), 31

Testes para cointegração, 172

VaR, 197

Variação regular no infinito, 70

Volatilidade realizada, 157

Volatilidade: fatos estilizados, 137

Referências Bibliográficas

- Accioly, V. B. e Mendes, B. V. M. Assessing the impact of the realized range on the (e)garch volatility: Evidence from brazil. *Brazilian Review of Econometrics*, (0):9–23, 2015.
- Aas, K., Czado, C., Frigessi, A. e Bakken, H. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182–198, 2009.
- Aas, K. e Berg, D. Models for construction of multivariate dependence - a comparison study. *European Journal of Finance*, 15(7-8):639–659, 2009.
- Aas, K. e Haff, I. H. The generalized hyperbolic skew student's t-distribution. *Journal of Financial Econometrics*, 4(2):275–309, Spring 2006. doi: 10.1093/jjfinec/nbj006.
- Abramowitz, M. e Stegun, I. *Handbook of Mathematical Functions*. National Institute of Standards and Technology, 1965.
- Aït-Sahalia, Y. How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies*, 18(2):351–416, 2005.
- Aït-Sahalia, Y., Mykland, P. A. e Zhang, L. Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics*, 160(1):160–175, January 2011.
- T. G. Andersen, T. Bollerslev, F. X. Diebold, and H. Ebens. The distribution of realized stock return volatility. *Journal of Financial Economics*, 61(1):43–76, July 2001a.
- T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys. The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96(453):42–55, 2001b. ISSN 01621459.
- Andersen, T. G. Some reflections on analysis of high-frequency data. *Journal of Business & Economic Statistics*, 18(2):146–153, 04 2000. doi: 10.2307/1392552. URL <http://www.jstor.org/stable/1392552>.
- Andersen, T. G., Bollerslev, T., Christoffersen, P. F. e Diebold, F. X. Chapter 15 volatility and correlation forecasting. *Handbook of Economic Forecasting*, 1:777–878, 2006.
- Andersen, T. G., Bollerslev, T., Diebold, F. X. e Ebens, H. The distribution of realized stock return volatility. *Journal of Financial Economics*, 61(1):43–76, 2001.

- Andersen, T. G., Bollerslev, T., Diebold, F. X. e Labys, P. The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96(453):42–55, 2001.
- Andersen, T. G., Bollerslev, T., Diebold, F. X. e Labys, P. Modeling and forecasting realized volatility. *Journal of the Econometric Society*, 71(2):579–625, 2003.
- Andersen, T. G., Bollerslev, T. e Diebold, F. X. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economic and Statistics*, 89(4):701–720, 2007.
- Andersen, T. G. e Bollerslev, T. Answering the skeptics: Yes standard volatility models do provide accurate forecasts. *International Economic Review*, 39(4):885–905, 1998a.
- Andersen, T. G. e Bollerslev, T. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39(4):885–905, 1998b.
- Anderson, C. W. e Dancy, G. P. *The Severity of Extreme Events*. University of Sheffield, Dpt. of Probability and Statistics, 1992.
- Arellano-Valle, R. B. , Branco, M. D. e Genton, M. G. A unified view on skewed distributions arising from selections. *The Canadian Journal of Statistics*, 34(4):121 – 121, 2006.
- Arellano-Valle, R. B., Gomez, H. W. e Quintana, F. A. A new class of sekew-normal distributions. Research report, PUC-Chile, 2003.
- Artzner, P., Delbaen, F., Eber, J. e Heath, D. Thinking coherently: Generalised scenarios rather than var should be used when calculating regulatory capital. *Risk-London-Risk Magazine Limited*, 10: 68–71, 1997.
- Artzner, P, Heath, D., Delbaen, F. e Eber, J.M. Coherent measures of risk. In *Documents from the 5th Annual Conference on Risk Management*. International Center for Business Information, Geneva, December 8th, 1998.
- Azzalini, A. A class of distributions which includes the normal ones. *Scandinavian J. Statistics*, 12: 171–178, 1985.
- Azzalini, A. Further results on a class of distributions which includes the normal ones. *Statistica*, 46: 199–208, 1986.
- Azzalini, A. e Capitanio, A. Statistical applications of the multivariate skew-normal distributions. *Journal of the Royal Statistical Society, Series B*, 61:579 – 602, 1999.
- Baillie, R. T., Bollerslev, T. e Mikkelsen, H. O. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74:3–30, 1996.
- Balke, N. S. e Fomby, T. Threshold cointegration. *International Economic Review*, 38(3):627–645, 1997.

- Bandi, F. M. e Russell, J. R. Microstructure noise, realized variance, and optimal sampling. *Review of Economic Studies*, 75(2):339–369, 2008. doi: 10.1111/j.1467-937X.2008.00474.x.
- Barndorff Nielsen, O. E. e Shephard, N. Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics*, 4(1):1–30, 2006a.
- Barndorff Nielsen, O. E. e Shephard, N. Impact of jumps on returns and realised variances: econometric analysis of time-deformed levy processes. *Journal of Econometrics*, 131(1-2):217–252, 2006b.
- O. E. Barndorff-Nielsen and N. Shephard. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal Of The Royal Statistical Society Series B*, 64(2): 253–280, 2002.
- O. E. Barndorff-Nielsen, P. R. Hansen, A. Lunde, and N. Shephard. Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica*, 76(6):1481–1536, 2008.
- O. E. Barndorff-Nielsen, P. R. Hansen, A. Lunde, and N. Shephard. Subsampling realised kernels. *Journal of Econometrics*, 160(1):204–219, January 2011.
- Barndorff-Nielsen, O. E. e Shephard, N. Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, 2(1):1–37, 2004.
- Barndorff-Nielsen, O. E. e Shephard, N. Estimating quadratic variation using realized variance. *Journal of Applied Econometrics*, 17(5):457–477, 2002.
- Beare, B. K. Copulas and temporal dependence. *Econometrica*, 78(1):395–410, 2010.
- Bedford, T. e Cooke, R. Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, 32:245–268, 2001. ISSN 1012-2443.
- Bedford, T. J. e Cooke, R. Vines - a new graphical model for dependent random variables. *Annals of Statistics*, 30(4):1031–1068, 2002.
- J. Beran. *Statistics for long-memory processes*. CRC Press, 1994.
- Berrendero, J., Mendes, B. V. M. e Tyler, D. E. On the maximum bias functions of MM-estimates and constrained M-estimates of regression. *Annals of Statistics*, 35(1):13–40, 2007.
- Bickel, P.J. e Doksum, K.A. *Mathematical statistics: basic ideas and selected topics*. Prentice Hall, 2001. URL <http://books.google.com.br/books?id=8poZAQAIAAJ>.
- Birnbaum, Z. W. Numerical tabulations of the distribution of kolmogorov’s statistic... *J. of American Statistical Association*, 47:425–441, 1952.
- Bollerslev, T. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.

- Bollerslev, T. A conditionally heteroskedastic time series model for speculative prices and rates of return. *The review of economics and statistics*, 69(3):542–547, 1987.
- Bollerslev, T. e Mikkelsen, H. O. Modeling and pricing long memory in stock market volatility. *Journal of Econometrics*, 73:151–184, 1996.
- Bollerslev, T. e Wooldridge, J. M. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, 11:143–172, 1992.
- Bortkiewicz, L. von. Variationsbreite und mittlerer fehler. *Sitzungsber. Berli. Math. Ges.*, 21:3 – 11, 1922.
- Box, G. E. P., Jenkins, G. M. e Reinsel, G. *Time Series Analysis, Forecasting, and Control*. Prentice-Hall 4th Edition, Englewood Cliffs, New Jersey, third edition, 1994.
- Box, G.E.P. e Pierce, D.A. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65:1509 – 1526, 1970.
- Branco, M.D. e Dey, D.K. A general class of multivariate skew-elliptical distributions. *J. Multivariate Analysis*, 79:93 – 113, 2001.
- Breymann, W., Dias, A. e Embrechts, P. Dependence structures for multivariate high-frequency data in finance. *Quantitative Finance*, 3(1):1–14, 2003. URL <http://econpapers.repec.org/RePEc:taf:quantf:v:3:y:2003:i:1:p:1-14>.
- Brockwell, P. J. e Davis, R. A. *Time series: theory and methods*. Springer-Verlag, 1991.
- Buckley, I., Saunders, D. e Seco, L. Portfolio optimization when asset returns have the gaussian mixture distribution. *European Journal of Operational Research*, 185(3):1434–1461, 3 2008. URL <http://www.sciencedirect.com/science/article/pii/S0377221706006242>.
- Campbell, J. Y., Lo, A. W. e MacKinley, A. C. *The Econometrics of Financial Markets*. Princeton, NJ: Princeton University Press, 1997.
- Chan, K. A further analysis of the lead-lag relationship between the cash market and stock index futures market. *Review of Financial Studies*, 5(1):123–52, 1992. URL <http://ideas.repec.org/a/oup/rfinst/v5y1992i1p123-52.html>.
- Chandra, M., Singpurwalla, N. D. e Stephens, M. A. Kolmogorov statistics for tests of fit for the extreme-value and weibull distributions. *Journal of the American Statistical Association*, 79(375): 729–731, 1981.
- Chen, X., Fan, Y. e Patton, A. J. Simple tests for models of dependence between multiple financial time series, with applications to u.s. equity returns and exchange rates. *SSRN eLibrary*, 2004. URL <http://ssrn.com/paper=513024>.
- Chow, Y. S. e Teicher, H. *Probability Theory*. Springer, 1988a.

- Chow, Y. S. e Teicher, H. *Probability theory: independence, interchangeability, martingales*. Springer-Verlag New York, 2a. edition, 1988b.
- Christensen, K. e Podolskij, M. Realized range-based estimation of integrated variance. *Journal of Econometrics*, 141(2):323–349, 2007.
- Christoffersen, P. e Pelletier, D. Backtesting value-at-risk: A duration-based approach. *Journal of Financial Econometrics*, 2(1):84–108, 2004.
- Coles, S. *An Introduction to Statistical Modelling of Extreme Values*. Springer Series in Statistics. Springer, 2001.
- Corsi, F. e Curci, G. A discrete sine transform approach for realized volatility measurement. *Manuscript, University of Pisa*, 2003.
- Czado, C. e Min, A. Bayesian inference for d-vines: Estimation and model selection. In *Dependence Modeling - Handbook on Vine Copulae*. World Scientific, 2009.
- Dacorogna, M., Müller, U., Pictel, O. e de Vries, C. The distribution of extremal foreign exchange rate returns in extremely large data sets. *Discussion paper, Tinbergen Institute*, 95(70), 1995.
- Danielsson, J. e De Vries, C.G. Value-at-risk and extreme returns. Research report, University of Iceland, 1997.
- Davidson, J. Moment and memory properties of linear conditional heteroscedasticity models, and a new model. *Journal of Business and Economic Statistics*, 22(1):16–29, 2004.
- Davis, R. A. e Mikosch, T. The sample autocorrelations of heavy-tailed processes with applications to arch. *The Annals of Statistics*, 26(5):2049–2080, Oct. 1998. URL <http://www.jstor.org/stable/120032>.
- Davis, R. A. e Resnick, S. I. Limit theory for bilinear processes with heavy-tailed noise. *The Annals of Applied Probability*, 6(4):1191–1210, Nov 1996. URL <http://www.jstor.org/stable/2245151>.
- de Haan, L. Sample extremes: an elementary introduction. *Statistica Neerlandica*, 30:161 – 172, 1976.
- de Haan, L. Slow variation and characterization of domains of attraction. *Statistical Extremes and Applications*, 131:31–48, 1984.
- de Haan, L., Resnick S. I., Rootzén, H. e de Vries, C. G. Extremal behaviour of solutions to a stochastic difference equation with applications to arch processes. *Elsevier*, 32(2):213–224, 1989.
- Dickey, D. A. e Fuller, W. A. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a):427–431, 1979.
- Diebold, F. X e Rudebusch, G. D. Long memory and persistence in aggregate output. *Journal of Monetary Economics*, 24(2):189–209, 1989.

- Dongming Zhua e John W. Galbraith. A generalized asymmetric student-t distribution with application to financial econometrics. *Journal of Econometrics*, 157:297 – 305, 2010.
- D. L. Donoho. Breakdown properties of multivariate location estimators. Ph.D. qualifying paper, Department of Statistics, Harvard University, 1982.
- Dupuis, D. J. e Field, C. A. Robust estimation of extremes. *The Canadian Journal of Statistics*, 26 (2):199–215, 1998.
- Durbin, J. Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables. *Econometrica*, 38(3):410–421, 1970.
- Efron, B. e Tibshirani, R. J. *An Introduction to the Bootstrap*. Chapman and Hall/CRC, 1993.
- Eitrheim, A. e Teräsvirta, T. Testing the adequacy of smooth transition autoregressive models. *Journal of Econometrics*, 74:59–76, 1996.
- Embrechts, P., Klüppelberg, C. e Mikosch, T. *Modelling Extremal Events for Insurance and Finance*, volume 12. Springer-Verlag, 1997.
- Embrechts, P., McNeil, A. J. e Straumann, D. Correlation and dependence in risk management: Properties and pitfalls. In M. Dempster, editor, *Value at Risk and Beyond*, chapter 7, pages 176 – 223. Cambridge University Press, 2002.
- Enders, W. e Falk, B. Confidence intervals for threshold autoregressive models. *Iowa State University, Dept. of Economics*, 1999.
- Enders, W. e Siklos, P. L. Cointegration and threshold adjustment. *Journal of Business and Economic Statistics*, 19(2):166–176, 2001.
- Engle, R. F. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, 1982.
- Engle, R. F. e Granger, C. W. J. Co-integration and error correction: representation, estimation and testing. *Journal of Business and Economic Statistics*, 55:251–276, 1987.
- Engle, R. F. e Lee, G. G. J. *A Permanent and Transitory Component Model of Stock Return Volatility*, volume 1, pages 475–497. Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W.J. Granger, oxford uk oxford university press edition, 1999.
- Engle, R. F. e Manganelli, S. Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, 22(2):367–381, 2004.
- Engle, R. F. e Patton, A. J. What good is a volatility model? *Quantitative finance*, 1(2):237–245, 2001.
- E. F. Fama. The behavior of stock-market prices. *Journal of Business*, 38(1):34–105, 1965.
- Fang, K., Kotz, S. e Ng, W. *Symmetric multivariate and related distributions*. Chapman & Hall, 1990.

- Fantazzini, D. Dynamic copula modelling for value at risk. *Frontiers in Finance and Economics*, 3: 705 – 730, 2007.
- Feller, W. *An Introduction to Probability Theory and Its Applications*. Wiley, NY, 1971.
- Fermanian, J. D., Radulovic, D. e Wegkamp, M. Weak convergence of empirical copula processes. *Bernoulli*, 10(5):847–860, 2004.
- Ferreira, C. A. Avaliação de modelos de risco através de backtesting. IMPA - Mestrado Profissionalizante. Orientadores: Dr. J. P. Zubelli e Dra. B. V. M. Mendes, 2013.
- Fielitz, B. D. e Rozelle, J. P. Stable distributions and the mixtures of distributions hypotheses for common stock returns. *Journal of the American Statistical Association*, 78(381):28–36, 03 1983.
- Fischer, M. J., Köck, C., Schlüter, S. e Weigert, F. Multivariate copula models at work: Outperforming the desert island copula? Discussion Papers 79, Friedrich-Alexander-University Erlangen-Nuremberg, Chair of Statistics and Econometrics, 2007.
- Fisher, R. A. e Tippett, L. H. C. Limiting forms of the frequency distributions of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, 24:180–190, 1928.
- Frahm, G., Junker, M. e Schmidt, R. Estimating the tail-dependence coefficient: Properties and pitfalls. *Mathematical Methods of Operations Research*, 55(2):301–327, 2004.
- Fréchet, M. Sur la loi de probabilité de l'écart maximum. *Ann. Soc. Polon. Math.*, 6(93), 1927.
- Frey, R. e McNeil, A. J. Estimation of tailed-related risk measures for heteroschedastic financial time series: an extreme value theory approach, department of mathematik, ethz. Technical report, Zurich, Working Paper, 1998.
- Fuller, W. A. *Properties of the Correction for Attenuation estimator*. Proceedings of the Social Statistics Section, 1976.
- Fung, W. K. H. e Hsieh, D. A. Empirical analysis of implied volatility: Stocks, bonds and currencies. Working papers, Fuqua School of Business, 1991.
- Garman, M. B. e Klass, M. J. On the estimation of security price volatilities from historical data. *The Journal of Business*, 53(1):67–78, January 1980.
- Genest, C., Quessy, J. F. e Rémillard, B. Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scandinavian Journal of Statistics*, 33(2):337–366, 2006. URL <http://ideas.repec.org/a/bla/scjsta/v33y2006i2p337-366.html>.
- M. Genton. *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*. Genton, M., 2004.
- Glosten, L. R., Jagannathan, R. e Runkle, D. E. On the relation between the expected value and the volatility of the nominal excess returns on stocks. *Journal of Finance*, 48(5):1779–1801, 1993.

- Gnedenko. B. V. . Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*, 44:423–453, 1943.
- Granger, C. W. J. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438, 1969.
- Granger, C. W. J. Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, 16(1):121–130, 1981.
- Granger, C. W. J. Developments in the study of cointegrated economic variables. *Oxford Bulletin of Economics and Statistics*, 48(3):213–228, 1986.
- Granger, C. W. J. e Joyeux, R. An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1(1):15–296, 1980.
- Granger, C. W. J. e Teräsvirta, T. *Modelling Non-linear Economic Relationships*. Oxford University Press, 1993.
- Greenwood, J. A., Landwehr, J. M. , MAtalas, N. C. e Wallis, J. R. Probability weighted moments: Definition and relation to parameters of several disitrbutions in inverse form. *Water Resources Research*, 15(5):1049–1054, 1979.
- Gumbel, E. J. Statistical theory of extreme values and some pratical applications. *Nat. Bureau of Standards Applications Mathematics Series*, 33:1–51, 1954.
- Gumbel, E. J. *Statistics of Extremes*. New York: Columbia University Press, 1958.
- Hall, P. Using the bootstrap to estimate mean square error and select smoothing parameters in non-parametric problems. *Journal of Multivariate Analysis*, 32:177–203, 1990.
- Hamilton, J. D. *Time Series Analysis*. Princeton Un. Press, January 1994. ISBN 0691042896.
- Hansen, B. Autoregressive conditional density estimation. *International Economic Review*, 35(3): 705–730, 1994.
- Hansen, P. R. e Lunde, A. Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, 24(2):127–161, 2006.
- Hendry, D. F. e Juselius, K. Explaining cointegration analysis: Part 1. *Energy Journal*, 21(1):1–42, 2000.
- Hill, B. A simple general approach to inference about the tail of a distribution. *The Annals of Mathematical Statistics*, 3:1163–1174, 1975.
- Hosking, J. R. M. L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society*, (52):105–124, 1990.
- Hosking, J. R. M. Fractional differencing. *Biometrika*, 68(1):165–176, 1981.

- Hosking, J. R. M. e Wallis, J. R. Parameter and quantile estimation for the generalized pareto disitribution. *Technometrics*, 29(3):339–349, 1987.
- Hosking, J. R. M. e Wallis, J. R. *Regional Frequency Analysis*. Cambridge University Press, 1997.
- Hosking, J. R. M. e Wallis, J. R. *Regional Frequency Analysis*. Cambridge University Press, 1997.
- Hosking, J. R. M., Wallis, J. R. e Wood, E. F. Estimation of the generalized extreme-value distribution by the method of probability weighted moments. *Technometrics*, pages 251–261, 1985.
- Hsing, T., Hüsler, J. e Leadbetter, M.R. On the exceedance point process for a stationary sequence. *Probability Theory and Related Fields*, 78:97–112, 1988.
- Huisman, R., Koedijk, K. G., Kool, C. J. M. e Palm, F. Tail-index estimates in small samples. *Journal of Business and Economic Statistics*, 19(1):208–216, 2001.
- Hurst, H. E. Long term storage capacity of reservoirs. *Trans. Am. Soc. Eng.*, 116:770–799, 1951.
- Hurvich, C. M. e Tsai, C. L. Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307, 1989.
- R. Ibragimov. Copula-based dependence characterizations and modeling for time series. Institute of Economic Research Working Papers 2094, Harvard University, 2005.
- Jansen, D. W. e de Vries, C. G. On the frequency of large stock returns: Putting booms and busts into perspective. *The REview of Economics and Statistics*, 73(1):18–24, 1991.
- Jansen, D.W. e De Vries, C.G. On the frequency of large stock returns: Putting booms and busts into perspective. *The Review of Economics and Statistics*, pages 18 – 24, 1991.
- Jarque, C.M. e Bera, A.K. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6(3):255–259, 1980.
- Jenkinson, A. F. . The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 87:158–171, 1955.
- Joe, H. *Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters*, volume 28, pages 120–141. 1996.
- Joe, H. *Multivariate Models and Dependence Concepts*. Chapman e Hall, London, 1997.
- Johansen, S. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12:231–254, 1988.
- Johansen, S. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, 1995.
- Johansen, S. e Juselius, K. Maximum likelihood estimation and inference on cointegration - with applications to the demand for money. *Oxford Bulletin of Economics and Statistics*, 52(2):169–210, 1990.

- Jondeau, E. e Rockinger, M. Testing for differences in the tails of stock-market returns. *Journal of Empirical Finance*, 10(5):559–581, 2003.
- Jones, M.C. e Faddy, M.J. A skew extension of the t -distribution, with applications. *J. R. Statist. Soc. B*, 65:159 – 174, 2003.
- Karatzas, I. e Shreve, S.E. *Brownian Motion and Stochastic Calculus*. New York: Springer-Verlag, 1988.
- Kon, S. J. Models of stock returns—a comparison. *The Journal of Finance*, 39(1):147–165, 03 1984.
- Kotz, S., Balakrishnan, N. e Johnson, N. *Continuous Multivariate Distributions*. New York: Wiley, 2000.
- Kupiec, P. H. Techniques for verifying the accuracy of risk measurement models. Finance and Economics Discussion Series 95-24, Board of Governors of the Federal Reserve System (U.S.), 1995.
- Kurowicka, D. e Cooke, R. M. Completion problem with partial correlation vines. *Linear Algebra and its Applications*, 418(1):188–200, 10 2006.
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. e Shin, Y. Testing the null hypothesis of stationarity against the alternative of a unit root : How sure are we that economic time series have a unit root? *Journal of Econometrics*, 54(1-3):159–178, 1992.
- Lamoureux, C. G. e Lastrapes, W. D. Heteroskedasticity in stock return data: Volume versus garch effects. *The Journal of Finance*, 45(1):221–229, 1990. ISSN 1540-6261.
- Leadbetter, M., Lindgren, G. e Rootzén, H. *Extremes and Related Properties of Random Sequences and Processes*. Berlin: Springer-Verlag, 1983.
- Lenum, M. O. A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics. *Oxford Bulletin of Economics and Statistics*, 54(3):461–472, 1992.
- Ljung, G.M. e Box, G.E.P. On a measure of a lack of fit in time series models. *Biometrika*, 65(2):297 – 303, 1978.
- Lo, A. Long-term memory in stock market prices. *Econometrica*, 59:1279–1313, 1991.
- Longin, F. The asymptotic distribution of extreme stock market returns. *Journal of Business*, 63:383 – 408, 1996.
- Lopuhaä, H. P. Multivariate Tau-estimators for location and scatter. *Canadian Journal of Statistics*, 19(3):307–321, 1991.
- Lopuhaä, H. P. Highly efficient estimators of multivariate location with high breakdown point. *Annals of Statistics*, 20(1):398–413, 1992.
- Lopuhaä, H. P. Asymptotics of reweighted estimators of multivariate location and scatter. *Annals of Statistics*, 27(5):1638–1665, 1999.

- Lopuhaä, H.P. e Rousseeuw, P.J. Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. *Annals of Statistics*, 19(1):229–248, 1991.
- Lukkonen, R., Saikkonen, P., and Teräsvirta, T. Smooth transition garch models. *Biometrika*, 75: 491–499, 1998.
- Maheu, J. M. e McCurdy, T. H. News arrival, jump dynamics, and volatility components for individual stock returns. *Journal of Finance*, 59(2):755–793, 2004.
- B. Mandelbrot. The variation of certain speculative prices. *Journal of Business*, 36(4):394–419, 1963.
- Mandelbrot, B. B. The stable paretian income distribution when the apparent exponent is near two. *International Economic Review*, 4(1):111–115, 1963.
- Mandelbrot, B. B. Computer experiments with fractional gaussian noises: Part 2, rescaled ranges and spectra. *Water resources research*, 5(1):242–259, 1969a.
- Mandelbrot, B. B. Computer experiments with fractional gaussian noises: Part 3, mathematical appendix. *Water Resources Research*, 5(1):260–267, 1969b.
- Mandelbrot, B. B. Broken line process derived as an approximation to fractional noise. *Water Resources Research*, 8(5):1354–1356, 1972.
- Martens, M. e van Dijk, D. Measuring volatility with the realized range. *Journal of Econometrics*, 138(1):181–207, 5 2007.
- McNeil, A. J. e Frey, R. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7:271–300, November 2000.
- McNeil, A. J., Frey, R. e Embrechts, P. *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton Series in Finance. Princeton University Press, New Jersey, 2005.
- McNeil, A.J. Calculating quantile risk measures for financial return series using extreme value theory. Technical report, Zurich: Department Mathematik ETH Zentrum, 1998. URL <http://dx.doi.org/10.3929/ethz-a-004320029>.
- Meddahi, N. A theoretical comparison between integrated and realized volatility. *Journal of Applied Econometrics*, 17(5):479–508, 2002.
- Mendes, B. V. M. *Introdução à Análise de Eventos Extremos*. Editora E-Papers, 2004. URL <http://www.e-papers.com.br>.
- Mendes, B. V. M. *Regressão Robusta: Conceitos, Aplicações, e Aspectos Computacionais*. Associação Brasileira de Estatística, São Paulo, 1999.
- Mendes, B. V. M. Assessing the bias of maximum likelihood estimates of contaminated garch models. *Journal of Statistical Computation and Simulation*, 67(4):359–379, 2000a.

- Mendes, B. V. M. Computing robust risk measures in emerging equity markets using extreme value theory. *Emerging Markets Quarterly*, 4(2):25–41, 2000b.
- Mendes, B. V. M. Asymmetric extreme interdependence in emerging equity markets. *Applied Stochastic Models in Business and Industry*, 21(6):483–498, 2005a. ISSN 1524-1904. doi: <http://dx.doi.org/10.1002/asmb.v21:6>.
- Mendes, B. V. M. Computing conditional var using time-varying copulas. *Brazilian Finance Review*, 3(2):249–263, 2005b.
- Mendes, B. V. M. e Accioly, V. B. . On the dependence structure of realized volatilities. *International Review of Financial Analysis*, 22(0):1 – 9, 2012. ISSN 1057-5219. doi: 10.1016/j.irfa.2012.01.001.
- Mendes, B. V. M. e Accioly, V. B. Robust pair-copula based forecasts of realized volatility. *Applied Stochastic Models in Business and Industry*, 2013. ISSN 1526-4025. doi: 10.1002/asmb.1960.
- Mendes, B. V. M. e Afube, C. Copula based models for serial dependence. *International Journal of Managerial Finance*, 7(1):68 – 82, 2011.
- Mendes, B. V. M. e Brandi, V. R. Modelling drawdowns and drawups in financial markets. *The Journal of Risk*, 6(3):53–69, 2004.
- Mendes, B. V. M. e Kolev, N. How long memory in volatility affects true dependence structure. *International Review of Financial Analysis*, 17(5):1070 – 1086, 2008. ISSN 1057-5219. doi: DOI: 10.1016/j.irfa.2007.06.008.
- Mendes, B. V. M. e Leal, R. P. C. Maximum drawdown: Models and applications. *Journal Of Alternative Investments*, pages 83–91, 2005.
- Mendes, B. V. M. e Leal, R. P. C. Portfolio management with semi-parametric bootstrapping. *Journal of Risk Management in Financial Institutions*, 3(2):174–183, 2010.
- Mendes, B. V. M. e Lopes, H. F. Data driven estimates for mixtures. *Computational Statistics and Data Analysis*, 47(3):583–598, 2004.
- Mendes, B. V. M. e Lopes, S. R. C. Bandwidth selection in classical and robust estimation of long memory. *International Journal of Statistics and Systems*, 1(1):167–190, 2006.
- Mendes, B. V. M. e Lopes, S. R. C. Dynamic copulas and long range dependence. *Frontiers in Finance and Economics*, 8(2):89–111, 2011.
- Mendes, B. V. M. e Marquez, D. Choosing an optimal investment strategy: The role of robust pair-copulas based portfolios. *Emerging Markets Review*, (13):449–464, 2012.
- Mendes, B. V. M. e Melo, E. F. Local estimation of copula based value-at-risk. *Brazilian Review of Finance*, 7:29–50, 2009.
- Mendes, B. V. M. e Melo, E. F. Local estimation of dynamic copula models. *International Journal of Theoretical & Applied Finance (IJTAF)*, 13:241–258, 2010.

- Mendes, B. V. M. e Souza, R. M. Measuring financial risks with copulas. *International Review of Financial Analysis*, 13(1):27–45, 2004.
- Mendes, B. V. M. e Tyler, D. E. *Constrained M-estimates for Regression*, volume 109, pages 299–320. Lecture Notes in Statistics: Robust Statistics, Data Analysis and Computer Intensive Methods, Springer Verlag, 1996.
- Mendes, B. V. M. e Tyler, D. E. Bias robustness of the CM-estimates of location and scale. *Brazilian Journal of Probability and Statistics*, 12(1):41–53, 1998.
- Mendes, B. V. M., Melo, E. F. L. e Nelsen, R. B. Robust fits for copula models. *Communication in Statistics*, 36:997–1017, 2007.
- Mendes, B. V. M., Semeraro, M. M. e Leal, R. P. C. Pair-copulas modeling in finance. *Financial Markets and Portfolio Management*, 24:193–213, 2010.
- Mittnik, S., Paolella, M. e Rachev, S. Unconditional and conditional distributional models for the nikkei index. *Asia-Pacific Financial Markets*, 5:99–128, 1998. ISSN 1387-2834. URL <http://dx.doi.org/10.1023/A:1010016831481>.
- Morettin, P. A. *Econometria Financeira*. Editora Blucher, 2008.
- Morettin, P. A. e Toloi, C. M. C. *Análise de Séries Temporais - 2ª Edição Revista e Ampliada*. Editora Edgar Blucher, 2006.
- Muler, N. e Yohai, V. Robust estimates for garch models. *Journal of Statistical Planning and Inference*, 138:2918—2940, 2008.
- S. Nefcti. *An introduction to the mathematics of financial derivatives*. California: Academic Press, 1996.
- Nelsen, R.B. *An introduction to copulas*. Springer, 2006.
- Nelson, D. B. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2): 347–370, 1991.
- Ng, S. e Perron, P. Unit root tests in arma models with data-dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association*, 90(429):268–281, 1995.
- Oakes, D. A model for association in in bivariate survival data. *J. Royal Statist. Soc. Ser. B (Stat. Method.)*, 44:414 – 422, 1982.
- Oliveira, T. Extremal distributions. *Revista da Faculdade Ciências A*, 3, 1958.
- Oliveira, T. Structure theory of bivariate extremes, extensions. *Est. Mat. Estat. e Econ.*, 7:165 – 195, 1962,1963.
- Parkinson, M. The extreme value method for estimating the variance of the rate of return. *The Journal of Business*, 53(1):pp. 61–65, 1980. ISSN 00219398.

- Patton, A. Modelling asymmetric exchange rate dependence. *International Economic Review*, 47(2): 527–556, 2006.
- Peiró, A. Skewness in financial returns. *Journal of Banking & Finance*, 23(6):847–862, 6 1999.
- Petrucelli, J. e Woolford, S. A threshold ar(1) model. *Journal of Applied Probability*, 21:270–286, 1984.
- Phillips, P. C. B. Understanding spurious regression in econometrics. *Journal of Econometrics*, 33(1): 311–340, 1986.
- Phillips, P. C. B. Optimal inference in cointegrated systems. *Econometrica*, 59(1):283–306, 1991.
- Phillips, P. C. B. e Ouliaris, S. Asymptotic properties of residual based tests for cointegration. *Econometrica*, 58(1):165–193, 1990.
- Pickands, J. III. Statistical inference using extreme order statistics. *Annals of Statistics*, 3:119–131, 1975a.
- Pickands, J. III. *Spline and Isotonic Estimation of the Pareto Function*, volume 12, pages 285–296. Statistical Extremes and Applications Proceedings of the NATO Symposium on Extreme Values, reindell publishers co. edition, 1984.
- Pickands, J. III. Statistical inference using extreme order statistics. *Annals of Statistics*, 3:119–131, 1975b.
- Reiss, R. D. *Statistical Analysis of Extreme Values*. Birkhauser Springer-Verlag, 1997.
- Reiss, R-D. e Thomas, M. *Statistical Analysis of Extreme Values*. Birkhäuser, Springer-Verlag, 1997.
- Remillard, B., Genest, C. e Beaudoin, D. Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44(2):199–213, April 2009. URL <http://ideas.repec.org/a/eee/insuma/v44y2009i2p199-213.html>.
- Rockinger, M. e Jondeau, E. Conditional dependency of financial series : an application of copulas. Les Cahiers de Recherche 723, HEC Paris, 2001.
- Roelant, E., Van Aelst, S. e Croux, C. Multivariate generalized s-estimators. *Journal of Multivariate Analysis*, 100(5):876–887, 2009.
- Rogers, L. C. G. e Satchell, S. E. Estimating variance from high, low and closing prices. *The Annals of Applied Probability*, 1(4):pp. 504–512, 1991. ISSN 10505164.
- Rosenblatt, M. Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, 23(3): 470–472, 1952.
- Rousseeuw, P. J. *A fast algorithm for the minimum covariance determinant estimator*, pages 283–297. W. Grossmann, G. Pflug, I. Vincze and W. Wertz. Reidel: Dordrecht, 1985.

- Rousseeuw, P. J. e Van Driessen, K. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41:212–223, 1999.
- Scott, D. W. *Multivariate density estimation: theory, practice, and visualization*, volume 275 of *Wiley series in probability and mathematical statistics*. Wiley, 1992.
- Sherman, B. Percentiles of the ω_n statistic. *Institute of Mathematical Statistics*, 28(1):259–261, 1957.
- Shih, J. H. e Louis, T. A. Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, 51(4):1384—1399, 1995.
- Sibuya, M. Bivariate extreme statistics. *Annals of the Institute of Statistical Mathematics*, 11:195 – 210, 1960.
- Sklar, A. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8:229–231, 1959.
- Smith, R. L. Estimating tails of probability distributions. *Annals of Statistics*, 15:1174–1207, 1987.
- Smith, R. L. Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1):67–90, 1985.
- Smith, R. L. Extreme value theory based on the r -largest annual events. *Journal of Hydrology*, 86 (1-2):27–43, 1986.
- Smith, R. L. e Weissman, I. Estimating the extremal index. *Journal of the Royal Statistical Society*, 56(3):515–528, 1994.
- Smith, R.L. Measuring risk with extreme value theory. Research report, University of North Carolina, Chapel Hill, 1999.
- Snedecor, G.W. e Cochran, W.G. *Statistical methods*. Number 276. Iowa State University Press, 8 edition, 1989.
- So, M. K. P. e Xu, R. Forecasting intraday volatility and value-at-risk with high-frequency data. *Asia-Pacific Financial Markets*, 20(1):83–111, 2013. ISSN 1387-2834. doi: 10.1007/s10690-012-9160-1.
- Stahel, W. A. *Robust estimation: Infinitesimal optimality and covariance matrix estimators*. PhD thesis, ETH, Zurich (in German), 1981.
- Stock, J. H. Asymptotic properties of least squares estimators of cointegrating vectors. *Econometrica*, 55(5):1035–1056, 1987.
- Tawn, J. An extreme value theory model for dependent observations. *J. Hydrology*, 101:227 – 250, 1988a.
- Tawn, J. Bivariate extreme value theory: Models and estimation. *Biometrika*, 77(22):245 – 253, 1988b.

- Ting, C. Which daily price is less noisy? *Financial Management*, 35(3), 2006.
- Tsay, R. S. *Analysis of Financial Time Series*. Wiley Series in Probability and Statistics. Wiley-Interscience, New Jersey, 2nd ed. edition, 2005.
- Tsay, R. S. e Tiao, G. C. Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary arma models. *Journal of the American Statistical Association*, 79(385):84–96, 1984.
- Van den Goorbergh, R. W. J., Genest, C. e Werker, B. J. M. Bivariate option pricing using dynamic copula models. *Insurance: Mathematics and Economics*, 37(1):101–114, 2005.
- Vandenhende, F. e Lambert, P. Local dependence estimation using semiparametric archimedean copulas. *Canadian Journal of Statistics*, 33(3):377–388, 2005.
- R. von Mises. *Mathematical Theory of Probability and Statistics*. New York, Academic Press, 1964.
- Wang, W. e Wells, M. T. Estimation of kendall's tau under censoring. *Statistica Sinica*, 10(1):1199–1215, 2000.
- Weissman, I. Estimation of parameters and quantiles based on the k largest observations. *J. of American Statistical Association*, 73:812–815, 1978.
- Wold, H. *A study in the analysis of stationary time series*. PhD thesis, Stockholm College, 1938.
- Yang, D. e Zhang, Q. Drift-independent volatility estimation based on high, low, open, and close prices. *The Journal of Business*, 73(3):477–91, July 2000.
- Yohai, V., Stahel, W.A., e Zamar, R.H. *A procedure for robust estimation and inference in linear regression*, pages 299–320. Directions in Robust Statistics and Diagnostics, Part II. Springer-Verlag, in stahel, w.a. and weisberg, s.w., eds. edition, 1991.
- Yohai, V.J. High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, 15(2):642–656, 06 1987.
- Zangari, P. Catering for an event. *Risk*, pages 34–36, 1997.
- L. Zhang, P. A. Mykland, and Y. Aït-Sahalia. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, 100(472):1394–1411, 2005.
- Zhang, L. Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli*, 12(6):1019–1043, December 2006. ISSN 13507265.
- Zhang, M. H. Modelling total tail dependence along diagonals. *Insurance: Mathematics and Economics*, 1(1):29–77, 2007.
- Zhou, B. High-frequency data and volatility in foreign-exchange rates. *Journal of Business & Economic Statistics*, 14(1):45–52, January 1996.
- Zivot, E. e Wang, J. *Modelling Financial Time Series with S-PLUS*. Springer, 2nd edition, 2006.

